THE UNIVERSITY *of* EDINBURGH

Sea Surface Temperature for Climate from the Along-Track Scanning Radiometers

Owen Embury

Doctor of Philosophy
University of Edinburgh
2013

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where listed below. This work has not been submitted in whole or part for any other degree.

Chapter 2 was originally published as O. Embury, C. J. Merchant, and M. J. Filipiak. A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Basis in radiative transfer. Remote Sensing of Environment, 116: 32-46, 2012.

- The introductory sections 2.1 and 2.2 were drafted by C. Merchant
- The work described in section 2.3.2 (Emissivity of seawater) was performed by M. Filipiak

Chapter 3 was originally published as O. Embury and C. J. Merchant. A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: A new retrieval scheme. Remote Sensing of Environment, 116: 47-61, 2012.

- The introductory section 3.1 was drafted by C. Merchant
- The discussion section 3.6 was drafted by C. Merchant

Chapter 4 was originally published as O. Embury, C. J. Merchant, and G. K. Corlett, A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Initial validation, accounting for skin and diurnal variability effects. Remote Sensing of Environment, 116:62-78, 2012.

- The introductory section 4.1 was drafted by C. Merchant
- G. Corlett maintains the ATSR match-up database described in section 4.2.2 and provided guidance on its contents.
- The Met Office skin and depth corrections described in section 4.3 were provided by K. Fennig.
- The discussion section 4.5 includes contributions from C. Merchant and G. Corlett.

Chapter 5 is intended for submission to Remote Sensing of Environment as: A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Harmonisation of Satellite Datasets into a single record.

(*Owen Embury*)

Abstract

This thesis describes the construction of a sea surface temperature (SST) dataset from Along-Track Scanning Radiometer (ATSR) observations suitable for climate applications. The algorithms presented here are now used at ESA for reprocessing of historical ATSR data and will be the basis of the retrieval used on the forthcoming SLSTR instrument on ESA's Sentinel-3 satellite.

In order to ensure independence of ATSR SSTs from in situ measurements, the retrieval uses physics-based methods through the use of radiative transfer (RT) simulations. The RT simulations are based on the Reference Forward Model line-by-line model linked to a new sea surface emissivity model which accounts for surface temperature, wind speed, viewing angle and salinity, and to a discrete ordinates scattering (DISORT) model to account for aerosol. An atmospheric profile dataset, based on full resolution ERA-40 numerical weather prediction (NWP) data, is defined and used as input to the RT model. Five atmospheric trace gases (N_2O, CH_4, HNO_3, and CFC-11 and CFC-12) are identified as having temporal and geographical variability which have a significant (~0.1 K) impact on RT simulations. Several additional trace gases neglected in previous studies are included using fixed profiles contributing ~0.04 K to RT simulations. Comparison against ATSR-2 and AATSR observations indicates that RT model biases are reduced from 0.2–0.5 K for previous studies to ~0.1 K.

A new coefficient-based SST retrieval scheme is developed from the RT simulations. Coefficients are banded by total column water vapour (TCWV) from NWP analyses reducing simulated regional biases to <0.1 K compared to ~0.2 K for global coefficients. An improved treatment of the instrument viewing geometry decreases simulated view-angle related biases from ~0.1 K to <0.005 K for the day-time dual-view retrieval. To eliminate inter-algorithm biases due to remaining RT model biases and uncertainty in the characterisation of the ATSR instruments the offset coefficient for each TCWV band is adjusted to match the results from a reference channel combination.

As infrared radiometers are sensitive to the skin SST while in situ buoys measure SST at some depth below the surface an adjustment for the skin effect and diurnal stratification is included. The same model allows adjustment for the differing time of observation between ATSR-2 and AATSR to prevent the diurnal cycle being aliased into the final record. The RT simulations are harmonised between sensors using a double-difference technique eliminating discontinuities in the final SST record.

Comparison against in situ drifting and tropical moored buoys shows the new SST dataset is

of high quality. Systematic differences between ATSR retrieved SST and in situ drifters show zonal, regional, TCWV, and wind speed biases are less than 0.1 K except for the most extreme cases (TCWV $<5\,\mathrm{kg\,m^{-2}}$). The precision of ATSR retrieved SSTs is ~0.15 K, lower than the precision of measurement of the global ensemble of in situ drifting buoys. From 1995 onwards the ARC SSTs are stable with instability of less than $5\,\mathrm{mK\,year^{-1}}$ to 95% confidence (demonstrated for tropical regions).

Lay Summary

The temperature of the sea surface plays an important role within the climate system so accurate knowledge of the sea surface temperature (SST) over recent decades is necessary both to quantify climate change and improve climate models. Existing SST records are based on broadly the same in situ ship and buoy data so any errors in the raw data will be common to all the analyses and will not be detected with inter-comparisons between the datasets. Therefore it is useful to have an independent record of SST which does not make use of in situ measurements. The Along-Track Scanning Radiometers (ATSRs) are a series of three instruments which have been carried on European Space Agency (ESA) satellites since 1991. They were specifically designed to allow estimation of SST independently of in situ measurements; however, they have not previously provided the accuracy or precision required for climate change research.

In this thesis I develop an improved method for estimating SST using the ATSR instruments. A physics-based model is used to simulate what the instruments should observe for a given set of atmospheric and surface conditions. The output of the physics-based "forward model" is then used to train an "inverse model" which predicts or retrieves the SST given a set of satellite observations. Compared to previous attempts to use ATSRs to measure SST I have improved both the forward model and the retrieval method.

The forward model is improved by using a more modern and complete simulation of the interaction of radiation with the surface and atmosphere. I have added gases, such as CFCs, which vary in concentration around the world and over time, but which were ignored or assumed constant in previous studies. The retrieval is improved by recognising that water vapour in the atmosphere is one of the primary factors biasing satellite retrievals of SST and correctly accounting of the angles at which the ATSR instruments observe the surface. Furthermore, satellites are sensitive to the temperature in the top 0.01 mm of the sea while ships and buoys can be measuring the temperature at a depth of 1 m or greater. Therefore models which account for the change of temperature with depth are used to produce SST measured at "satellite", "buoy", and "ship" depths.

Comparisons against in situ measurements show that the new ATSR SST retrievals are highly accurate with regional biases relative to buoys less than 0.1 K. From 1995 onwards the new ATSR SST retrievals can measure changes in SST to an accuracy of 5 mK per year. This will be an important, independent, check that traditional buoy and ship based records of SST are correctly measuring SST

changes over recent decades.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

The temperature of the sea surface plays a fundamental role within the Earth's climate system; ocean covers over 70% of the Earth's surface and constitutes the bulk of the thermal inertia of the climate system (e.g. Hegerl and Bindoff, 2005). Global sea surface temperature (SST) is therefore a major component of the observational record used to assess climate change (Trenberth et al., 2007) and it is important to have trustworthy data. When estimating subtle changes in SST over a number of years the SST measurement system must be stable. This is a major concern for historical (1850 to present) global SST, as the instrument types and technologies used to create in situ-based historical records have changed significantly over time (Kent et al., 2010; Kennedy et al., 2011b), raising the question about whether the changes in technology (see Fig. 1.1) are affecting the observed climate changes. One way to build confidence that the observed changes are real, and not artefacts of the SST measurement system, would be to compare them against changes estimated from an independent dataset (e.g. Immler et al., 2010). Historical analyses by different groups of researchers are largely based on the same ship and buoy measurements (Merchant et al., 2008c) so they cannot be considered independent of each other.

In recent decades, satellite retrievals of SST have become available, but these have generally used methods which tie them empirically to in situ observations (Kilpatrick et al., 2001; Merchant and Le Borgne, 2004). The series of Along Track Scanning Radiometer (ATSR) instruments were specifically designed to provide accurate satellite-based retrievals of SST independently of in situ measurements by use of radiative transfer (RT) simulations to define the retrieval algorithm (Závody et al., 1995). Operationally produced ATSR SST retrievals were recognised as highly accurate space based records (Corlett et al., 2006) and were used to bias correct other datasets (e.g. Le Borgne et al., 2012; Blackmore et al., 2012; O'Carroll et al., 2012; Donlon et al., 2012b). However, the ATSR SST retrievals used were neither accurate nor stable enough for climate-change studies (Merchant et al., 2008c).

This thesis presents a new ATSR retrieval algorithm suitable for generating data relevant for climate

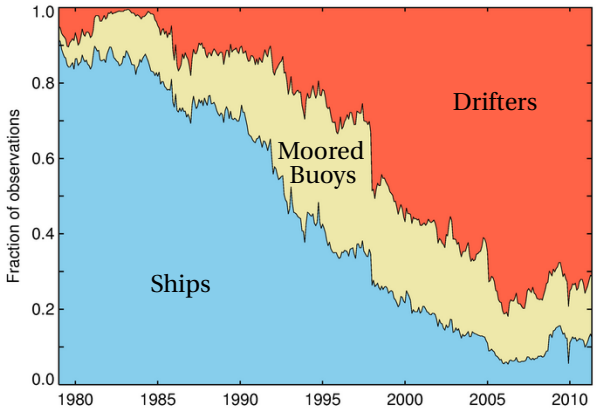

Figure 1.1: Distribution of in situ types present in historical record since 1980 (Image supplied by J. Kennedy).

applications which is used to produce a 20 year time series of SST independent of in situ observations. The research presented herein represents part of my contribution to the ATSR Reprocessing for Climate (ARC) project described in (Merchant et al., 2008c).

1.2 Layout of thesis

The remainder of this introductory section discusses the importance of SST in the Earth's climate, how SST is measured, and the specific requirements for SST data used for climate-change studies. The subsequent chapters comprise a series of papers discussing the construction of the ARC SST dataset. Chapter 2 (Embury et al., 2012a) covers the radiative transfer simulations which provide the physical basis of the ARC SST retrieval independent of in situ data. Chapter 3 (Embury and Merchant, 2012) develops an improved SST retrieval coefficient design which greatly reduces systematic biases from geometric and atmospheric effects compared to previous attempts. Chapter 4 (Embury et al., 2012b) presents an initial comparison of the ARC SST retrievals against in situ drifters and tropical moored buoys accounting for the different measurement depths. Chapter 5 (Embury and Merchant 2013) explains the harmonisation technique used to combine data from all three instruments into a single consistent dataset. This thesis concludes with chapter 6 which includes a summary of the independent validation of the ARC SSTs and future work.

1.3 Importance of SST

Sea surface temperature is a key geophysical variable in the Earth's climate and accurate knowledge of SST is vital for climate studies, ocean forecasting, and weather prediction. Figure 1.2 shows a simple summary of the Earth's global mean energy balance (Kiehl and Trenberth, 1997; Le Treut et al., 2007). About one third of the solar energy which reaches the Earth is reflected back in to space by clouds and the surface. Approximately 30% of the remainder is absorbed by the atmosphere while 70% is

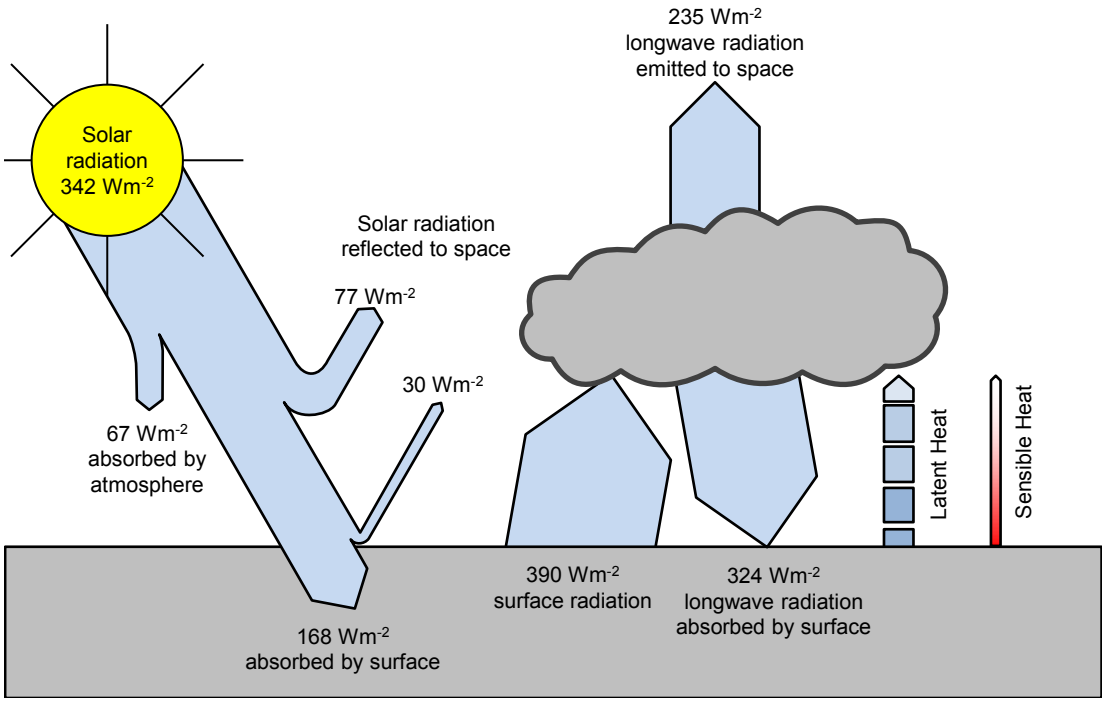

Figure 1.2: Summary of Earth's mean global radiation budget with flux estimates from Kiehl and Trenberth (1997)

absorbed by the surface. As ocean covers over 70% of the Earth's surface and is less reflective than the land the oceans are absorbing over half the solar energy entering the climate system.

Heat transfer from the surface to the atmosphere comprises three parts: net infrared radiation, latent heat, and sensible heat. Net infrared radiation is the difference between the thermal radiation emitted by the surface and the downwelling thermal radiation emitted by the atmosphere. Latent heat is heat transferred by the evaporation and condensation of water. Sensible heat is transferred through conduction and convection. Overall the net heat transfer from the surface to the atmosphere is approximately double the solar energy absorbed by the atmosphere; as the largest source of energy input to the atmosphere, surface fluxes are therefore an important driver of weather systems. Surface fluxes are driven by the temperature difference between the surface and atmosphere, so surface temperature, including SST, is a key input to Numerical Weather Prediction (NWP) models.

The oceans act as an important reservoir of heat, absorbing solar energy while the sun is shining and continuously releasing heat into the atmosphere. Variability in heat flux is present on sub-daily to seasonal timescales. During the day, the ocean absorbs approximately half the incident solar radiation in the top 2 m while some visible wavelengths can penetrate down to 20 m deep depending on water clarity (Fairall et al., 1996). Under conditions of sustained low wind speed this can result in the formation of a warm layer of surface water which is stable under low wind conditions as it will be less dense than the underlying water; under windy conditions waves will mix the surface water so the warm layer does not form. At night, the water cools from the surface as heat is lost to the atmosphere.

This drives convection in the upper ocean which, along with wind-driven mixing, will mix water throughout the upper ocean known as the ocean mixed layer. The mixed layer will gradually warm throughout the spring and summer months while the solar energy absorbed is greater than the heat loss to the atmosphere, then cool during the autumn and winter when the solar input decreases.

The thickness of the ocean mixed layer can vary from 10 to 200 m (de Boyer Montégut et al., 2004) which means the ocean has a much higher thermal inertia than the land, where diurnal heating extends less than a metre below the surface. Below 1 m depth, the ground temperature is effectively constant throughout the year. This occurs because the land is solid, so heat can only be transported below the surface by conduction; there is no process to mix the heat lower in the ground and virtually all the heat absorbed during the day will be released at night. Consequently continental climates have greater temperature variation with hotter summers and colder winters than maritime climates.

The ocean mixed layer also has a far greater heat capacity than the atmosphere. This is primarily due to the greater density of water than air. The mass per unit area of the atmosphere can be approximated as p_0/g where $p_0 = 101325$ Pa is mean sea level pressure and $g = 9.81$ m s^{-1} is the acceleration due to gravity; giving a total atmospheric mass per square metre of 10,329 kg m^{-2}. As the density of sea water is just over 1000 kg m^{-3} the atmosphere has the same mass as approximately 10 m depth of water. At the same time the specific heat capacity (energy required to heat 1 kg of a substance by 1 K) is approximately 4 times greater for sea water than air. Therefore the heat capacity of the atmosphere is equivalent to 2.5 m of ocean.

1.4 Measuring SST

1.4.1 SST definitions

Before discussing techniques for measuring SST it is useful to consider what is meant by the term "sea surface temperature", as different methods of measuring it will produce different results. The fundamental problem is that "surface" is poorly defined in the context of measuring SST. Different measurements of SST are actually measurements of the temperature at a finite depth below the surface; the depth of measurement for "SST" typically varying from ~10 µm to ~10 m. The temperature differential across the upper 10 m of the ocean can range from tenths of a kelvin to several kelvin depending on heat fluxes and ocean turbulence. Figure 1.3 shows two idealised temperature profiles for the upper ocean during day and night, there are two main features which describe the profiles: the cool-skin and warm-layer.

The cool-skin or skin effect refers to the steep temperature gradient which exists in the top ~1 mm of the sea. This layer is known as the "skin layer" within which heat is primarily transported via molecular conduction since turbulent processes are suppressed by the air-sea boundary. The temperature decreases as we approach the air-sea interface as the net flux of heat is typically from the ocean to the atmosphere. Sensible and latent heat transfer both occur at the air-sea interface while

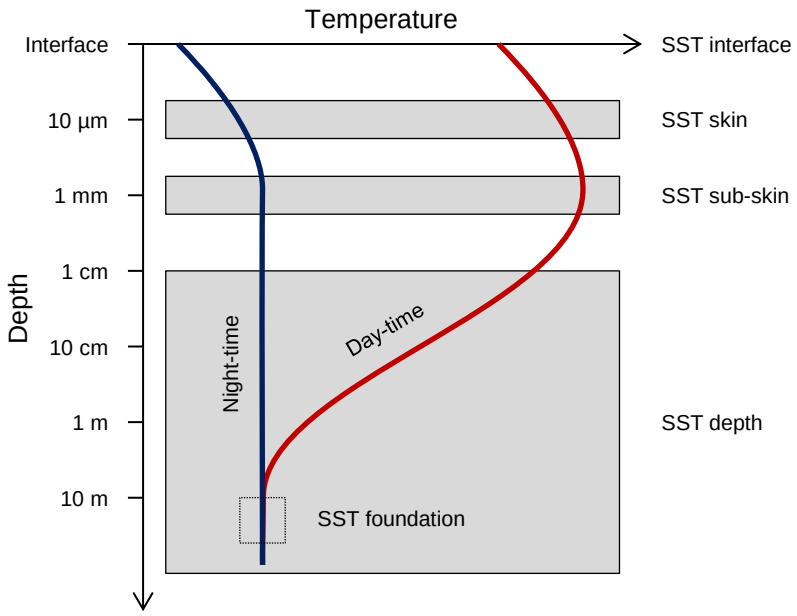

Figure 1.3: Hypothetical vertical temperature profile of upper 10 m of ocean for a well mixed night-time case (blue line) and thermally stratified day-time case (red line).

longwave radiation has a penetration depth of 10–20 µm; hence heat is being lost from the top of the skin layer. The cool skin is typically 0.1–0.5 K (Fairall et al., 1996) and is almost always such that the skin is cooler than the water below; however, it is possible for it to be reversed if the net non-solar heat flux is into the ocean.

As mentioned in the previous section, the upper ocean is heated by solar radiation during the day. About half the solar energy is infrared or near-infrared and is absorbed within the first metre or two of the ocean, the remainder is visible wavelengths which may penetrate down to 20 m depth. Typically surface winds will mix the heat lower into the ocean so that the surface warming is just a few tenths of kelvin; but under sustained low wind conditions a warm-layer can form which can be over 5 K warmer than the pre-dawn temperature (Gentemann et al., 2008; Merchant et al., 2008a).

In order to reduce the ambiguity associated with the term "SST" the Group for High Resolution Sea Surface Temperature (GHRSST Donlon et al., 2007) has proposed a framework of five different "types" of SST which provide a link between what can be defined and what can be measured. These are the interface, skin, subskin, depth, and foundation and are illustrated in Fig. 1.3 with typical temperature profiles for day and night.

Interface Temperature (SSTint)

The interface temperature represents the temperature of the sea at the actual interface between the sea and the air. It is the interface temperature which determines the latent and sensible heat fluxes, as well as gas exchange across the boundary. However, the interface temperature is not directly measured.

Skin Temperature (SSTskin)

Skin temperature is defined as the temperature measured by an infrared radiometer operating in the 3–12 μm wavelength range (the majority of infrared satellite instruments use bands centred around 3, 11, and 12 μm). The penetration depth of radiation into the ocean depends on wavelength and, for these infrared frequencies, is approximately 10–20 μm which is near the top of the skin-layer.

SST-skin is the correct SST to use for calculating the upwelling longwave flux from the ocean as it is this emitted thermal radiation which an infrared radiometer is measuring. SST-skin is also the most appropriate measured SST to use for calculating the sensible and latent heat fluxes (Fairall et al., 2003) as it is the closest measurement to SST-interface.

Sub-skin Temperature (SSTsubskin)

The sub-skin temperature is the temperature measured by a microwave radiometer operating in the 6–11 GHz frequency range. The exact depth will depend on frequency, physical conditions, and viewing geometry, but is of order 1 mm. As such it is below the skin-layer so provides an adequate representation of temperature at lower depths during night when the upper ocean is well mixed.

Depth Temperature (SSTdepth)

This definition covers any measurement made by physical contact with the ocean — i.e. the majority of in situ measurements other than ship based radiometers. The typical measurement depth is ~1 cm to 10 m; although profiling floats and thermistor chains can measure at significantly greater depths these are no longer considered to be surface temperature measurements.

During the night when the ocean is well mixed SSTdepth is more-or-less constant with depth, but during the day there can be a significant vertical structure due to solar heating of the ocean surface (both day and night situations are significantly complicated if there is heat flux from the lower ocean). As such, the GHRSST science team recommend that all SSTdepth measurements specify the depth at which the measurement was made. In practice this meta-data is not always available; for instance, the depth of drifting buoy measurements depends on both the type of buoy and the prevailing conditions at the time of measurement, while historical measurements from ships may depend on the length of rope available.

Foundation Temperature (SSTfnd)

The foundation SST is a hypothetical temperature representing the temperature free of diurnal variability, or the temperature at the first time of day when the heat gain from solar input exceeds the heat loss at the surface. Foundation SST was adopted as a replacement for the historical concept of “bulk SST” typically used to represent the temperature of the ocean mixed layer and assumed independent of depth throughout the layer.

1.4.2 Ship based measurements

Usable ship-based measurements of SST date back to the 1850s when the then sea-faring nations agreed on various standardised measurement techniques at the Brussels Maritime Conference of 1853 (Rayner et al., 2006; Maury, 1858, 1859). Therefore, with over 150 years of data, ship-based measurements are the longest running record of SST available. However, they also have the highest degree of variability; not all aspects of SST measurement were standardised in 1853 and they have varied both from country to country and over time as technologies have developed.

Most ships making SST observations are Voluntary Observing Ships (VOS) which are supplied with standardised equipment and training for its use by their national weather service. VOS data are therefore concentrated along the major shipping lanes which have changed over the years. The number of VOS vessels peaked in the mid 1980s and has been declining since then.

Other ship-based sources of SST includes Research Vessels (RV) and weather ships. RVs are dedicated to marine research and can supply data away from shipping lanes. They will take a wide range of oceanographic measurements depending on the primary aims of their research campaign. Weather ships were stationary ships used to provide meteorological observations between the late 1940s to mid 1970s when automated buoys came into use (Rayner et al., 2006). The final weather ship was decommissioned on 1st January 2010 (Schiermeier, 2009).

Bucket

Early measurements of SST were made using a bucket to obtain a sample of sea water from the ocean and measuring its temperature on the ship's deck. Bucket measurements were commonly used until sometime between the 1920s and 1960s depending on country (Rayner et al., 2006). Bucket measurements will sample the near-surface water, but the water can cool while is is being hauled back up to the deck. The rate of heat loss will depend on a variety of factors: the height of the ship, the speed the ship is travelling, the size of the bucket, and the material used to make the bucket.

Engine intake

Some ships will measure the temperature of the engine intake water, used to cool the engines. The depth at which the water is taken will depend on how heavily the ship is loaded, but is likely to be several metres below the surface. Engine intake measurements are often biased warm as the water may pass through long runs of pipes and pumps before it is measured and the engine itself is a source of heat (Kennedy et al., 2011b).

Hull contact

In recent years dedicated hull contact sensors have been installed on some ships. These are expected to be more accurate and less noisy than engine intake measurements, but they are still relatively untested (Kennedy et al., 2011b).

Radiometers

Radiometers have been deployed on some ships to measure the infrared radiation emitted by the ocean skin-layer with accuracies of 0.1 K or better (Donlon et al., 2008). However, they are expensive and delicate instruments which require an operator to maintain them and protect them during rain or poor weather. Therefore most radiometers have only been deployed on RVs. One instrument, the Infrared SST Automonous Radiometer (ISAR) was developed as a fully automated system which has been deployed on VOS vessels (Donlon et al., 2008).

As radiometers are sensitive to SST-skin rather than SST-depth they are useful for validating satellite retrievals of SST (e.g. Wimmer et al., 2012; Kearns et al., 2000); however, they are limited by the low volume of data available. Other uses of radiometers include investigating the skin-effect (Donlon et al., 2002; Minnett et al., 2001) and the spectral emissivity of the sea surface (Minnett et al., 2001).

1.4.3 Moored Buoys

Moored buoys are floating platforms, typically discs, with diameters from a couple of metres up to 12 m. The buoys are anchored to the sea bed with a cable and carry a range of meteorological instruments including SST sensors. Moored buoys can be separated into two distinct groups: coastal moorings and the Global Tropical Moored Buoy Array (GTMBA). Coastal moorings are mostly located off the coasts of the United States and Europe. They are deployed and managed by national buoy programmes; different countries often using different types of buoys so the depth of SST measurement varies.

The GTMBA includes tropical buoys managed by three different programmes: TAO/TRITON (Tropical Atmosphere Ocean/Triangle Trans Ocean Buoy Network; McPhaden (1995)), PIRATA (Prediction and Research Moored Array in the Atlantic; Bourlès et al. (2008)), and RAMA (Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction McPhaden et al. (2009)). All three components of the GTMBA use Autonomous Temperature Line Acquisition System (ATLAS) moorings which measure a wide range of meteorological and sub surface parameters, including SST at a depth of 1 m. The instruments measure SST at a resolution of 0.001 K and accuracy of 0.02 K every 10 minutes. The buoys transmit daily average and hourly average SSTs while a communication satellite is overhead (hence hourly measurements are not transmitted for all 24 hours in the day). However, the full resolution data is recorded on the buoys and recovered when the buoys are serviced.

1.4.4 Drifting Buoys

Drifting buoys were originally developed for measuring ocean currents, but they are now the largest source of in situ SST measurements as all modern drifters include a SST sensor. Development of modern drifting buoys began in the late 1970s with the advent of satellite positioning systems (Lumpkin and Pazos, 2007). While initial drifter designs were highly variable, work towards a standardised

drifter design began in 1982 as part of the World Climate Research Program. The first modern Surface Velocity Program (SVP) drifters were deployed in 1993 with an updated "mini" SVP introduced in 2002.

A modern drifter comprises a spherical surface float between 30 cm and 40 cm diameter tethered to a subsurface drogue centred at 15 m below the surface (Lumpkin and Pazos, 2007). The float includes a temperature sensor near the bottom which measures SST at a nominal depth of 20 cm; however the exact depth will vary with surface wave conditions and the entire float may be temporarily submerged. Floats will have either a submergence sensor or a strain sensor to detect if the drogue has become detached. Some floats may include extra instruments to measure surface pressure, wind speed, salinity, or ocean colour etc. Data is transmitted using an on board radio via the Argos satellite tracking system which determines the buoys location.

The Global Drifter Network developed rapidly through the 1990s and reached "full" coverage of 1250 buoys in 2005 and is maintained by deployments of new buoys managed by the Global Drifter Program (Lumpkin and Pazos, 2007). Today drifters provide the most complete in situ coverage of the world; however, as they follow ocean currents the coverage is far from uniform as drifters will collect in ocean convergence zones. There is particularly poor coverage in the North Pacific, Arctic, Southern Ocean, and upwelling regions such as the Tropical Warm Pool. O'Carroll et al. (2008) estimate that global drifter SSTs are accurate to ~0.23 K.

1.4.5 Satellites

Satellite based radiometers are used to estimate SST by measuring the thermally emitted surface radiation (Fig. 1.2) which reaches space. Since most surface-emitted radiation is absorbed by the atmosphere, the instruments must operate at wavelengths where the atmosphere is mostly transparent and use two or more channels in order to account for the residual atmospheric signal. There are usable atmospheric windows in the infrared and microwave parts of the spectrum. Infrared radiometers are more accurate and have higher spatial resolution, but they are limited to cloud free regions and are more sensitive to variable clear-sky atmospheric conditions. Microwave radiometers can sense the surface through clouds, but have lower spatial resolution, are less accurate, and are sensitive to surface roughness and rain.

Typical infrared radiometers have three channels centred near wavelengths of 3.7 µm, 11 µm, and 12 µm. Each channel has a different sensitivity to water vapour: the 3.7 µm is the least sensitive meaning that nearly all the measured signal is surface emitted radiance, while the 12 µm is the most sensitive to water vapour and over half of the measured signal may come from the atmosphere. As a result the signal measured in each channel will be a different combination of surface and atmospheric contributions. The surface temperature can then be retrieved using a weighted combination of two or more different channels as described in chapter 3.

Passive microwave radiometers used for measuring SST have channels between 7 GHz and 37 GHz. At these frequencies non-precipitating clouds are largely transparent and atmospheric water vapour

has a negligible impact on observations at the lower end of the frequency range. However, microwave radiometers cannot "see through" rain and they are affected by radio frequency interference from TV broadcasts and other communications. The main limitations of microwave radiometers are their lower resolution and lower accuracy compared to infrared instruments. The spatial resolution of microwave instruments is restricted by the aperture of the radio telescope used, with the resolution inversely proportional to the aperture size. The Advanced Scanning Microwave Radiometer (AMSR) has a footprint of ~60 km and a 1.6 m aperture [AMSR ATBD]. Increasing the resolution to ~30 km would require a 3.2 m telescope or synthetic aperture techniques. Microwave radiometers are less accurate than infrared radiometers because the Earth's thermal emission is much less at microwave wavelengths than infrared (thermal emission from the Earth peaks around 10 µm).

Traditionally, SST retrievals are defined empirically using collocated satellite observations and in situ measurements of SST (e.g. Kilpatrick et al., 2001). As such, any biases or errors in the in situ data used to train the retrieval are folded into the retrieval itself. Furthermore the satellite is not sensitive to the same SST as measured by the in situ sensor. However, this is not always accounted for in the retrieval.

A major advantage of instruments carried on polar orbit satellites is that they will image the whole of the Earth every 1 to 3 days. Even a single infrared instrument, limited to clear-sky observations, has higher spatial coverage than the entire in situ network. In 2003, about 25% of 5°longitude-latitude cells in the Hadley Centre SST analyses did not have at least one in situ SST reported each calendar month while they did have ATSR observations (Merchant et al., 2008c). Conversely, polar orbiting satellite based instruments have a lower temporal sampling rate than in situ.

Polar orbiting satellites used for remote-sensing are often placed in a sun-synchronous orbits. In a sun-synchronous orbit the orbital plane precesses at the same rate that the Earth orbits around the sun, i.e. at 360°per year, such that the satellite will always overpass a given latitude at the roughly the same local solar time (ascending and descending overpasses are separated by 12 hours). This is ideal for many remote sensing applications as the sun remains in the same place in the sky when a given point on the surface is observed. An important attribute of a sun-synchronous orbit is its Local Equatorial Crossing Time (LECT); for example, the Envisat satellite has a LECT of 1000 will overpass the equator at 10am and 10pm local time.

1.5 Climate records

For climatological studies of the ocean we need to quantify marine changes over decadal time periods or longer. This requires the ability to detect a small trend signal of order 0.1–0.2 K decade^{-1} (Trenberth et al., 2007) in the presence of far larger seasonal (~10 K) and diurnal signals (which can be over 5 K Gentemann et al., 2008; Merchant et al., 2008a). In order to achieve this it is necessary to have an SST record which is stable, i.e. constant in the absence of real SST changes and not incorrectly showing changes in the observing system as changes in SST.

Stability is a key concern for historical records of SST due to changes in both measurement technologies and the sampling of the ocean over time (Kent et al., 2010). Measurements of SST have been made from ships for over 200 years although there were not done so regularly until the 1850s (see subsection 1.4.2). Shipping routes have changed over this time due to social and political reasons so that different parts of the ocean were observed over the years. At the same time the technologies used have changed from wooden buckets, to canvas buckets, to rubber buckets, to engine intake, to hull contact thermometers. Buoys have been used to measure SST since the 1970s and in recent decades have provided the bulk of in situ SST observations as the number of ship measurements has declined. Unlike ships, buoys are either in a fixed location (moored) or follow ocean currents (drifters). Each of these changes results in slightly different estimates of the SST so that the changes in the observing system can distort our estimate of the climate signal. Much of the work in generating historical SST analyses such as HadSST (Kennedy et al., 2011b) is to account for the changes in instrumentation and ensure the analyses are stable.

Ideally, we need an independent SST record which can be used to verify that the changes observed in the historical SST analyses are real changes and not artefacts of the evolving measurement techniques. When two independent datasets based on different measurement principles give the same estimate of change we will have increased confidence that the changes are real and not systematic errors inherent in the datasets (e.g. Immler et al., 2010). The use of different measurement techniques is key to this to ensure that the two datasets are not affected by the same "unknown" systematic errors. For instance, historical ship-based records of SST could be separated by country of origin of ship, but even though records from one country are (in one sense) independent of another, they are all subject to similar changes in technology and measurement techniques over time, and so are likely to be affected by the same systematic errors.

Satellite-derived records of SST are available for the last three decades and have been blended into some analyses of historical SST (e.g. Reynolds et al., 2002; Rayner et al., 2003) or used to produce modern satellite based records such as the Advanced Very High Resolution Radiometer (AVHRR) Pathfinder dataset (Casey et al., 2010). However, these satellite products are empirically tied to the in situ measurements (see subsection 1.4.5); while they are effectively "filling in the gaps" between in situ measurements they are not adding an independent estimate of SST.

Creating a stable and independent SST record from ATSR data is feasible but had not been fully achieved prior to this present work. The ATSR instruments, described in more detail in the next section, were designed to allow satellite-based estimation of SST without the use of in situ data. Such a dataset is useful not only as an independent verification of the in situ record during the overlap between the two datasets, but can also benefit the long-term datasets prior to the satellite era in two ways (Merchant et al., 2008c). Firstly, it can be used to characterise biases and uncertainties in modern in situ measurements (e.g. Kennedy et al., 2012). These may be applicable to earlier decades and can help improve estimates of bias and uncertainty in the historical records. Secondly, the data can be used to improve understanding of the spatial and temporal modes of SST variability which are used

to reconstruct spatially complete SST fields when observations are sparse (e.g. Kaplan et al., 1998; Rayner et al., 2003; Smith and Reynolds, 2004).

1.6 ATSR

The Along Track Scanning Radiometers (ATSRs) are a series of three instruments which have been carried on European Space Agency (ESA) satellites since 1991. The first, ATSR-1, was part of the payload on the European Remote Sensing (ERS)-1 satellite launched in July 1992; ATSR-2 was carried on the ERS-2 launched in April 1995; the final Advanced ATSR (AATSR) was onboard the Envisat satellite launched in March 2002. The ATSRs were intended from the outset to produce accurate and stable SST data independently of any in situ observations. For this they have several design features which make them suitable for climate applications unlike sensors such as the AVHRRs which were intended for meteorological applications. Full details of the ATSR instrument design can be found in Edwards et al. (1990), Stricker et al. (1995), and Smith et al. (2012) for the ATSR-1, ATSR-2, and AATSR respectively. The key design features are briefly discussed below.

Firstly, the instrument spectral response functions (SRF) were accurately measured during pre-flight calibration and characterisation (see Mason, 1991; Smith et al., 1993; Smith, 1999; Smith et al., 2001). This allows us to produce accurate simulations of satellite data using a forward or Radiative Transfer (RT) model which are used as the basis of the SST retrieval (Závody et al., 1995; Merchant et al., 1999; Embury et al., 2012a). As such the errors inherent in ATSR retrievals of SST come from the RT model and not in situ data.

Secondly, the ATSRs were exceptionally well calibrated; they used a two-point calibration system specified to drift less than 0.03 K over the sensor lifetime (Mason et al., 1996; Smith et al., 2001). The use of two on-board calibration targets allows the blackbodies to be held at temperatures of 260 K and 300 K which greatly reduced detector non-linearity effects over the range of scene temperatures used for SST retrieval. Other instruments typically use a single on-board target held around 290 K and rely on a view of space (2.7 K) to provide the second point.

Third, the solid-state detectors that measured the infrared radiance were actively cooled to approximately 82 K to reduce instrument noise and avoid temperature dependent effects on the calibration. The instrument fore-optics were also cooled below ambient temperature reducing the impact of instrument self-emission and the variation of the self-emission due to temperature cycling. Along with cooling, the instruments are enclosed and include baffles to prevent stray-light affecting the detectors.

Next, the ATSR instruments had a unique dual-view capability. The main telescope used an inclined conical scan pattern giving it two views of Earth's surface: a nadir view, and a forward view (~55°). Having two views of the surface allows the instrument to gather more information and more effectively separate surface and atmospheric effects; i.e. the SST retrieval can be made more robust to atmospheric conditions, including water vapour and stratospheric aerosol.

Finally the ERS and Envisat satellites which carried the ATSRs were all stable platforms with near-constant LECT throughout the life of the platforms. The ERS-1 and ERS-2 platforms had LECTs of 10:30 and Envisat had a crossing time of 10:00 all of which were maintained within a few minutes. For climate data, the half-hour difference between ERS-2 and Envisat must be accounted for, but this is much simpler than the case for instruments on drifting platforms such as AVHRR (e.g. Jin and Treadon, 2003) or the Microwave Sounding Unit.

As a result, the ATSR instruments have produced the most accurate satellite record of SST currently available (Corlett et al., 2006; O'Carroll et al., 2006; Wimmer et al., 2012). Work by O'Carroll et al. (2008) using a three-way error analysis showed that AATSR data were more precise (standard deviation of error 0.16 K) than in situ drifter measurements (0.23 K); Kennedy et al. (2012) use AATSR data as a reference for assessing the accuracy of in situ SST measurements. ATSR data are often used as a reference to bias correct other satellite datasets including: the Spinning Enhanced Visible and Infra-Red Imager (SEVIRI; Le Borgne et al., 2012), the AVHRRs (Blackmore et al., 2012; O'Carroll et al., 2012), the Visible Inrared Spin Scan Radiometer (VISSR; Merchant et al., 2003), and the Met Office Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system (Donlon et al., 2012b).

However, the existing ATSR SST retrievals are neither accurate nor stable enough for climate-change studies. Operational ATSR SST retrievals are still affected by regional systematic biases (Merchant et al., 2006b) up to ~0.3 K (O'Carroll et al., 2006; Embury et al., 2012b). There are inconsistencies between sensors of 0.1–0.4 K depending on channel combination (Merchant et al., 2008c; Veal et al., 2013). Comparison of ATSR-1 SSTs to Tropical Atmospheric Ocean array moorings showed a drift of 0.7 K decade^{-1} (Merchant et al. 2008) while O'Carroll et al. (2006) found a drift of ~0.2 K decade^{-1} in ATSR-2 SSTs relative to buoy SSTs. As such, the ATSR SST retrievals still need substantial improvement to bring maximum benefit to climate research studies.

1.7 ATSR Reprocessing for Climate

The research presented in this thesis is part of my contribution to the ATSR Reprocessing for Climate (ARC) project described in Merchant et al. (2008c). The aim of the ARC project was to produce a time series of SST from the ATSRs for climate research and involved researchers at the University of Edinburgh, the University of Leicester, the UK Met Office, the National Oceanography Centre, and the Science and Technology Facilities Council. Targets for the project were chosen by colleagues at the Hadley Centre for Climate Prediction and Research by considering the question: "what would be the characteristics of a satellite SST record that would bring significant advance in climate change research?" Their chosen requirements were:

1. A minimum record length of 15 years
2. A high degree of independence of in situ records
3. Stability better than 5 mK year^{-1}

4. Biases <0.1 K regionally (i.e. scales of a few thousand kilometres).

5. Estimates of both skin and depth SST.

6. Comprehensive error characterisation.

In order to meet these stringent targets, new ARC processing software was developed to allow different SST retrieval and cloud screening algorithms to be compared by applying them to the complete archive of Level 1b ATSR data (calibrated, geolocated, top-of-atmosphere brightness temperature and reflectance) held at the NERC Earth Observation Data Centre (NEODC). The ARC software was designed to be efficient and flexible allowing the entire input dataset of approx 30 TB (compressed) to be reprocessed producing outputs from all the different streams (different combinations of algorithms) in a couple of weeks. By developing and testing all the new science and algorithms in a production environment from the start I ensured that the project outcomes were of immediate practical use and would not be affected by any "scaling up" problems.

The cloud screening algorithms considered were the operational cloud detection scheme described in Závody et al. (2000) and a new probabilistic Bayesian scheme developed within the project based on Merchant et al. (2005) with updates for day-time observations described in Mackie et al. (2010). The Bayesian cloud detection was found to be more effective (better at detecting cloud and less prone to incorrectly flagging clear observations as cloudy) and was used for the final ARC outputs. In addition a test for Saharan dust contamination Good et al. (2012) developed at the University of Leicester which built upon an earlier Saharan dust detection for SEVIRI technique (Merchant et al., 2006a) was added to prevent biases associated with dust events.

Three SST retrieval techniques were compared: the operational (prior to this work) retrieval algorithm as described in Závody et al. (1995) and Merchant et al. (2006a); a new coefficient based retrieval method (this work); and an experimental optimal estimation retrieval (e.g. Merchant et al., 2008b). The new coefficient based SST retrieval was found to be the most accurate, precise, and consistent and its development is documented in this thesis.

To ensure that the final ARC products were of suitable quality they were validated by teams at University of Leicester, the UK Met Office, and the National Oceanography Centre. This "external"validation will be briefly discussed in chapter 6.

The remainder of this thesis documents the new ARC SST retrieval method. Chapter 2 describes the radiative transfer simulations which underpin the retrieval providing a physical basis independent of in situ SST observations. Chapter 3 develops the SST retrieval coefficient design, addressing the shortcomings of the previous operational algorithm. Chapter 4 presents my validation of the retrievals against in situ data. Chapter 5 explains the harmonisation technique used to create a single consistent dataset from the three instruments.

Chapter 2

Basis in Radiative Transfer

Owen Embury, Christopher J. Merchant, Mark J. Filipiak

Abstract

We present new radiative transfer simulations to support determination of sea surface temperature (SST) from Along Track Scanning Radiometer (ATSR) imagery. The simulations are to be used within the ATSR Reprocessing for Climate project. The simulations are based on the "Reference Forward Model" line-by-line model linked with a sea surface emissivity model that accounts for wind speed and temperature, and with a discrete ordinates scattering model (DISORT). Input to the forward model is a revised atmospheric profile dataset, based on full resolution ERA-40, with a wider range of high-latitude profiles to address known retrieval biases in those regions. Analysis of the radiative impacts of atmospheric trace gases shows that geographical and temporal variation of N_2O, CH_4, HNO_3, and CFC-11 and CFC-12 have effects of order 0.05, 0.2, 0.1 K on the 3.7, 11, 12 μm channels respectively. In addition several trace gases, neglected in previous studies, are included using fixed profiles contributing ~0.04 K to top-of-atmosphere BTs. Comparison against observations for ATSR2 and AATSR indicates that forward model biases have been reduced from 0.2 to 0.5 K for previous simulations to ~0.1 K.

2.1 Introduction

The temperature of the sea surface is a parameter of fundamental importance within Earth's climate, since the upper ocean constitutes the principal thermal inertia of the climate system (e.g. Hegerl and Bindoff, 2005). The series of Along Track Scanning Radiometers (ATSRs) are instruments that, by design, are capable in principle of extremely accurate determination of sea surface temperature (SST) from space - as described in the introduction to this special issue (Llewellyn-Jones et al, this issue). From the start it was intended that ATSR SSTs should be obtained independently of in situ

measurements, by use of radiative transfer simulation to define ATSR retrieval coefficients (Závody et al., 1995). In this paper we present radiative transfer simulations to be used as the basis of a new linear SST retrieval for the ATSR instruments. This includes both the radiative transfer model, and the set of atmospheric and surface states used as input.

The initial attempts at SST estimation using coefficients based on radiative transfer (RT) were hampered by two factors.

First, despite the ATSR's dual-view capability being well designed for the task, there was lack of clarity about how to estimate SST in the presence of stratospheric aerosol (e.g. Brown et al., 1997). Significant loadings of volcanic aerosol were present in the stratosphere following the 1991 eruption of Mount Pinatubo (e.g. Baran and Foot, 1994), throughout the commissioning phase and first two years of operations of ATSR.

Second, understanding of the spectroscopy of water vapour in the window regions of the infrared spectrum, used for SST observation, was inadequate: in particular, the parameterisation of the water vapour continuum absorption (Clough et al., 1989) available to Závody et al. (1995) was associated with biases in brightness temperature simulations of up to several tenths of a kelvin (Merchant et al., 1999, see Fig. 5 therein). As pointed out by Merchant and Le Borgne (2004), biases in simulated brightness temperature cause a bias in estimated SST of $\mathbf{a}^{\mathrm{T}}\varepsilon$ where the column vector $\mathbf{a}$ lists the weights given to each channel used in SST retrieval, and the column vector ε contains the bias in the RT simulations of BTs for those channels in the same order. Since in general $|\mathbf{a}| > 1$, RT biases tend to be amplified into larger biases in the retrieved SST, and in consequence RT simulations need to be accurate to < 0.1 K for RT-based coefficients to give SSTs with biases ~ 0.1 K.

These two problematic factors were ameliorated by recognition that retrieval coefficients could be designed to be orthogonal ("robust") to the effects of stratospheric aerosol, and by use of a revised parameterisation of the water vapour continuum absorption (Merchant et al., 1999; Merchant and Harris, 1999). Subsequently, there have been further improvements in spectroscopic data (Rothman et al., 2003; Clough et al., 2005).

This present article describes the basis in radiative transfer for SST retrieval in the context of the ATSR Reprocessing for Climate (ARC). ARC's objective is to create and exploit a time series of SST from the ATSRs for climate research. Our targets are that SST biases should be reduced to 0.1 K on seasonal-regional scales, and that the time series should be homogeneous and stable, such that trend artefacts should be no greater than 0.05 K decade^{-1}. To do this, the ARC consortium are integrating new developments in cloud detection (e.g. Merchant et al., 2005) and improved understanding of SST retrieval errors (e.g. Merchant et al., 2006b) with consistent and rigorous use of improved RT modelling. This latter aspect is described here. The motivation, objectives and overall framework of ARC are reported in more detail in Merchant et al. (2008c).

2.2 Sea surface temperature estimation

The principles of SST estimation by use of linear or near-linear coefficients applied to infrared BTs have been well rehearsed in the literature (e.g., Anding and Kauth, 1970; Deschamps and Phulpin, 1980; Merchant and Le Borgne, 2004). Here we very briefly set the context of SST retrieval for the rest of this paper.

Most commonly, the SST estimate is a linear or near-linear combination of observed brightness temperatures:

$$\hat{x} = a_0 + \mathbf{a}^{\mathrm{T}}\mathbf{y} \tag{2.1}$$

where $\hat{x}$ is the estimated SST, a_0 is an offset coefficient, and the elements of $\mathbf{a}$ are the retrieval coefficients (weights) that multiply in turn the BTs in the observation vector $\mathbf{y}$.

In some ways this approach is deceptively simple, in that subtle systematic errors get wrapped into the estimated SST (Merchant et al., 2009). These tend to be somewhat reduced if multiple sets of coefficients are defined for more restricted ranges of conditions (e.g Závody et al., 1995; Shenoi, 1999). The ultimate extension of this is to define coefficients dynamically for each situation observed, which in effect is optimal estimation (Merchant et al., 2008b). This requires a fast RT model (which is inevitably more approximate than a full model). This approach is explored within ARC and will be reported elsewhere. The focus for this paper is detailed forward modelling in support of determination of coefficients for SST retrieval.

The term "forward modelling" here describes use of a radiative transfer model (RTM) to simulate satellite observations and includes the step of selecting the set of atmospheric and surface states to be simulated. In the next section, we describe the components of the RTM used. We then describe the profiles of atmospheric temperature, humidity, gas composition and aerosol composition defined for simulation. The final section of the paper presents some comparisons of observed with simulated distributions of brightness temperature and draws some conclusions about the forward modelling undertaken. The derivation of coefficients and their initial validation against in situ observations are presented in companion papers (Embury and Merchant, 2012; Embury et al., 2012b).

2.3 Components of the Radiative Transfer Model

The RTM comprises three components. There is the gas transmittance model, which takes in atmospheric profiles and fundamental spectroscopic data, calculates the transmittance through the atmosphere layer-by-layer and line-by-line, and integrates the radiative transfer equation at high spectral resolution. Coupled to that is a model that defines the surface spectral emissivity, with inputs related to the sea surface state. Aerosol effects are more complicated because of three-dimensional scattering in addition to absorption, and proved impractical to simulate at high spectral resolution. Aerosol effects are therefore estimated by passing channel-integrated outputs of the transmittance model to a discrete ordinates solver. We now describe these components in turn.

2.3.1 Atmospheric transmittance model

Two line-by-line layer-by-layer atmospheric transmittance models were considered in detail: the "Reference Forward Model" (RFM; Dudhia, 2006), and the LBLRTM (Clough et al., 2005).

RFM is based on the GENLN2 program (Edwards, 1992) and was developed at the University of Oxford to be used as a reference forward model for the MIPAS instrument. RFM includes version 1.1 of the "MT_CKD" water vapour continuum parameterisation (Clough et al., 2005), which is identical to the more recent version 1.3 of the parameterisation in the spectral regions observed by the ATSR instruments.

LBLRTM stands for "line-by-line radiative transfer model" and we use version 9.4. In terms of modelled spectroscopic processes it is equivalent to RFM except that it adds CO_2 P/R branch line mixing in the v2 region (600–800 cm^{-1}). Nevertheless LBLRTM was significantly faster, 5 to 10 times, than RFM.

Brightness temperatures simulated by the two models were comparable to levels of order ±0.02 K when applied to identical profiles. A review of these same two models as a reference model for fast RTM parameterisation (Matricardi, 2007) concluded that the differences between the two models was much lower than the difference between the models and measurements. We chose to use the RFM, the model with which we had longer experience.

We use the TES spectroscopic database (version 1.3) — maintained by AER Inc. — based on HITRAN2000 (Rothman et al., 2003) with all official updates through to August 2003 and additional updates (to CH_4, CO, O_3, HOOCH, O_2) which were later included in HITRAN2004 (Rothman et al., 2005). The TES line parameters are used in preference to the more recent HITRAN edition in order to maintain necessary consistency with the water vapour continuum model used (Clough et al., 2005).

Atmospheric layering is dependent on the input profiles described in section 2.4. The temperature and water vapour profiles (ERA-40) are on 60 model levels from surface to approximately 56 km; other gases (MIPAS reference profiles) are on a fixed altitude grid with 1 km spacing at the surface increasing to 5 km at top of profile (120 km). The model levels are used below 60 km and the fixed altitude grid used above (temperature and water vapour data taken from the reference MIPAS atmosphere) giving a total of 72 levels. When integrating the radiative transfer equation we assume a spherical Earth for determining the local path length and use the linear in tau approximation for calculating the radiance emitted from each layer.

Line by line calculations were performed at a spectral resolution of 0.01 cm^{-1}, calculating (when coupled to the spectral surface emission model, subsection 2.3.2) both top of atmosphere radiance spectra and transmittance spectra for each atmospheric layer. Radiance spectra are then convolved with the instrument spectral response functions (SRFs) to give channel-integrated radiance. These in turn are related to BT by interpolation of a look-up table derived by convolving ideal Planck emission with the SRFs. The look-up table introduces a maximum error of ~0.001 K in the 3.7 µm channel and less than half this for the 11 and 12 µm channels. Note that the conversion of observed radiances to

BTs is likewise based on interpolation of a look-up table.

2.3.2 Calculation of emissivity of seawater

Values of the infrared emissivity of water surfaces accounting for view angle, wind speed, salinity, and temperature dependence are required for thermal remote sensing of the ocean and lakes. For ARC, a new tabulation of emissivity has been generated following a review of data and methods of calculation. A compilation of the refractive indices of seawater and pure water between 500 and 3500 cm^{-1} was made at 274, 287, and 300 K. The refractive indices were then used to calculate emissivity as function of wave number (600–3350 cm^{-1}), view angle (0–23° and 51–57°), temperature (270–310 K), and wind speed (0–25 $\mathrm{m\,s}^{-1}$ at 12.5 m). Preliminary studies had previously demonstrated that improved emissivity, including its temperature dependence, is an important part of achieving the goals of the ARC project (Merchant et al., 2008c). Our new synthesis of refractive indices and emissivity builds on the results of Newman et al. (2005).

The emissivity at any view angle is modelled assuming the sea surface is made up of plane facets with a given slope distribution (which depends on wind speed), calculating the Fresnel reflection coefficients for each facet, and summing the contribution of each facet (Masuda et al., 1988). In addition to the direct emission, the contribution from emitted radiation that has been reflected by the surface into the view angle is included (Watts et al., 1996; Wu and Smith, 1997). The important parameters for the emissivity model are the sea slope distribution and the refractive indices.

Slope distribution

The model uses the clean surface slope distribution measured/modelled by Cox and Munk (1954). The isotropic Gaussian version of the distribution is used, with mean slope zero and variance (or mean squared slope)

$$\sigma^2 = 0.003 + 0.00512u \tag{2.2}$$

where u is the wind speed ($\mathrm{m\,s}^{-1}$) at 12.5 m height.

Since Cox and Munk (1954) there have several measurements of slopes, models of wave spectra to derive the mean square slope, and reanalyses of Cox and Munk's data (see references in, e.g., Gatebe et al. (2005), Hwang (1997), and Liu et al. (2000)). However, there is no compelling case to choose any of these over Cox and Munk's distribution. Laboratory measurements are difficult to scale to the open ocean (Hwang, 1997) and results at radar frequencies, e.g. Wentz (1977), will underestimate the mean squared slope applicable for IR and visible wavelengths (Hwang, 1997; Liu et al., 2000). The directional dependence in several sets of measurements is close to Cox and Munk's results (Hwang, 1997). It is probable that at low wind speeds the mean squared slope is a background value due to swell: removing this effect gives good agreement between all the visible-wavelength measurements, confirming the linear dependence of mean squared slope on wind speed (Hwang, 1997). The constant term in Eq. 2.2

is retained in the current emissivity model to give at least some estimate of the background mean squared slope due to swell.

The normal distribution with mean squared slope given by Eq. 2.2 is an approximation. The slope distribution is slightly non-normal and has directional dependence (the ratio of the along wind to total is approximately 0.6). In addition, the measurements were made in neutral or stable conditions. Using the current emissivity model, the effect of ignoring the directional dependence was a ±0.001 uncertainty in the emissivity at a view angle of 55° (this is equivalent to a wind speed error of 0.2 m s^{-1}). The effect of ignoring the non-normality and stability was not calculated.

Refractive indices of water

The refractive indices of pure water are based on the measurements of Newman et al. (2005) in the frequency range 760–1230 cm^{-1} and the compilation of Downing and Williams (1975) elsewhere. A more recent compilation (Bertie and Lan, 1996) gives results very different from Downing and Williams (1975) near 800 cm^{-1} and 3400 cm^{-1}. Bertie and Lan (1996) use the Attenuated Total Reflection technique while the results of Downing and Williams (1975) results are from reflection and transmission measurements so it is likely that there are actual, systematic differences which make the choice between the data sets difficult. Newman et al. (2005) found that their emissivity data (from an instrument similar to the ATSRs) fitted the Downing and Williams (1975) data much better, thus Downing and Williams (1975) seems more appropriate for the calculation of emissivity for ATSRs, at least near 800 cm^{-1}.

The temperature dependence of the refractive indices is taken from Newman et al. (2005) in the range 760–1230 cm^{-1}, set to zero in the range 1230–2000 cm^{-1} (outside the ATSR channel range), and taken from Pinkley et al. (1977) elsewhere. The salinity dependence is taken from Pinkley and Williams (1976) (correcting the plotting error in n near 900 cm^{-1} in their Figure 5), who have measurements at all wavenumbers, rather than Friedman (1969), who uses a constant correction at all frequencies above 1200 cm^{-1}. The temperature dependence and salinity dependence were assumed independent and added together to get the refractive indices for sea water (35 PSU).

The errors are estimated as the root sum of squares of the error bars in Downing and Williams (1975) and Pinkley and Williams (1976), and the quantization error (0.005) in Pinkley et al. (1977). Near the peaks at 600 and 3400 cm^{-1} the errors have been increased to 0.03, as suggested by Bertie and Lan (1996).

Emissivity calculations

We calculate emissivity as function of wave number at various wind speeds, temperatures, and view angle. Results for the direct emissivity calculation match those of Masuda et al. (1988). The calculation including the surface emitted surface reflected (SESR) contribution is done having corrected the error in equation 23 of Wu and Smith (1997) to give a smooth transition with angle. Nadir emissivities at

the peak near 900 cm^{-1} are 0.992 for 300 K sea water, which compares well with Figure 2 of Newman et al. (2005). Temperature dependence in the 11–12 μm region is similar to that shown in Figure 7 of Newman et al. (2005). Salinity dependence is similar to Figure 8 of Newman et al. (2005). Wind dependence is similar to Wu and Smith (1997) and the more recent model of Masuda (2006): increasing wind gives decreasing emissivity at 55°. For example, at 910 cm^{-1}, 300 K, 55°, emissivity at 15 $m\,s^{-1}$ minus emissivity at 0 $m\,s^{-1}$ equals: -0.0029 for the current model, -0.0035 in Figure 6 of Wu and Smith (1997), and ~-0.002 in Masuda (2006).

The calculated emissivity does not include any effect of foam, which will affect the results at higher wind speeds. The maximum effect is described in Watts et al. (1996), but measured foam emissivities are likely to be smaller (see Figure 5 of Salisbury et al. (1993)) and the temperature of the foam may not be the skin temperature (active foam appears warmer and residual foam cooler then the surrounding sea surface (Marmorino and Smith, 2005)).

2.3.3 Aerosol and scattering model

For scattering calculations we used the discrete ordinates radiative transfer model DISORT. As it is not practical to apply this model to monochromatic spectra we use channel integrated transmissivities to calculate the effects due to scattering. This is done by running DISORT for both the aerosol affected and clear-sky profiles. These BT differences, referred to here as ΔBT, are then added to the clear-sky BTs calculated by integrating the monochromatic radiances.

Optical properties of marine aerosols are taken from the Optical Properties of Aerosols and Clouds (OPAC; Hess et al., 1998) database. This dataset includes both microphysical and optical properties for ten different aerosol components; the optical properties calculated from the microphysical properties using Mie theory. These basic components can be combined to represent the various mixtures observed in the atmosphere (i.e. maritime, continental etc).

Although the OPAC dataset does include a sulfate component, it is primarily intended to represent sulfate particles found in Antarctic aerosol rather than the stratosphere. Furthermore as only one size distribution is available it cannot be used to represent the evolution of the stratospheric aerosol after the Pinatubo erruption in 1991. Optical properties for stratospheric sulfate were instead calculated using Mie scattering theory and H_2SO_4 refractive index data (Tisdale et al., 1998) assuming a lognormal size distribution. Measurements from Deshler et al. (2003) were used for various size distributions representing the evolution of stratospheric aerosol present during the ATSR lifespans. Full details on the choice of size distributions are given in section 2.4.3.

Finally, we use the optical properties of Saharan dust from Highwood et al. (2003) in preference to the OPAC mineral-transported component. Highwood et al. used AERONET observations from Dakar, Senegal to derive the aerosol optical properties. When modelling the impacts of Saharan dust on the MSG SST retrieval (Merchant et al., 2006a) we found that the Highwood parameters matched observations more closely than the OPAC mineral-transported component [see also Good et al., this

issue].

2.4 Atmospheric and surface state

Three categories of state variables need to be defined for simulation. First, there is the meteorological state. This is represented by profiles of temperature and humidity, and by the mean sea level pressure, surface temperature and surface wind speed. Simulations are only for clear-sky conditions since retrievals are only done for clear-sky areas, and therefore cloud parameters are not relevant. Second, there is the composition of radiatively active gases other than water vapour - what we refer to as "trace"gases (although there is also some relevant absorption by the continuum absorption of the dominant atmospheric gases N_2 and O_2). Such gases are not considered to vary with meteorological state (although in reality, some do so in geographically complex ways). To capture the most important variability of such gases for climate purposes, trace gas compositions are chosen to vary, according to need, with respect to season, location and long-term trend. Third, there is aerosol, as characterized by a vertical profile of single-scatter albedo, extinction and asymmetry parameter. The approach to each of these categories of variable is described next.

2.4.1 Meteorological state

The original radiative transfer calculations for ATSR performed by Závody et al. (1995) used a set of 160 radiosondes to represent the set of atmospheric states encountered globally. More recently, output from numerical weather prediction (NWP) models has been used (Merchant et al., 1999) as they provide a more representative distribution of the atmospheric states encountered. Chevallier (2002) documents a set of 13495 atmospheric profiles sampled from the ECMWF 40 year re-analysis project (ERA-40) that represent the full variability of the complete dataset. The dataset includes profile data of temperature, water vapour, ozone, and surface fields such as temperature and wind speed. These profiles were filtered using the following criteria, following Merchant et al. (1999), to form the basis of the ARC profile set:

- All land or mixed land/ocean profiles excluded
- Profiles in locations of >95% sea ice excluded
- Profiles with any layers where relative humidity exceeds 95% excluded

This resulted in a core set of 2194 profiles; however there were very few high latitude profiles, especially in summer. Therefore the core profile set was supplemented with 668 profiles extracted from ERA-40 to ensure uniform latitudinal and seasonal variation.

The profile set as described so far includes profiles which represent conditions in the presence of sea-ice where the average surface temperature in the grid cell is below 271.35 K. Merchant et al. (1999) excluded these profiles from the calculations as the resulting BTs are clearly not suitable for SST

retrievals. However, this means that high latitude profiles with potentially large air-sea temperature differences were not represented, which is thought to be responsible in part for the large biases seen in current ATSR SST retrievals at high latitude. For ARC we retain these profiles, approximately 60% of profiles polewards of 60°, but replace the surface temperature, representing the average of ice and sea water in the ~100 km cell, with temperatures chosen according to an empirical model (Matthiesen and Merchant, 2003):

$$SST = T_1 + (1 - ice)(T_0 - T_1)\frac{n_1}{5} + \frac{n_2}{5} \tag{2.3}$$

where: $T_0 = 2\,^{\circ}\mathrm{C}$ and $T_1 = -1.8\,^{\circ}\mathrm{C}$; ice is the sea ice coverage (0 to 1); and n_1 and n_2 are random variables sampled from a Poisson distribution with mean of 5. This simple model was chosen to represent the general features of the relation between sea temperature and ice coverage seen in HadISST (Rayner et al., 2003).

2.4.2 Trace gases

Overview

While water vapour is usually the most significant absorber which affects satellite observations in the window channels used by the ATSRs, other gases still affect the observed BTs and must be accounted for. Previously, such gases have been assumed to have fixed concentrations (Závody et al., 1995) with no geographical or temporal variation; this, however, can lead to associated geographical biases and, in the case of CO_2 or CFCs which have long term trends, long term drifts in bias.

In order to identify which gases should be included, RT simulations were performed where only water vapour absorption was considered, and a two component atmosphere including water vapour and the gas to be considered. In each case the MIPAS mid-latitude daytime model atmosphere (Remedios et al., 2007) was used. The final gas list which includes all gases with an impact greater than 0.001 K on simulated BTs is shown in Table 2.1. Figure 2.1 shows the spectral signatures for these gases along with the ATSR spectral response functions.

The effect of geographical variation in gas concentrations are shown in Table 2.2, here the same two component atmosphere has been used except all five MIPAS model atmospheres (equatorial, mid latitude day/night, polar summer/winter) are used for the gas concentrations. Gases which had a regional impact less than 0.001 K are not shown.

Carbon dioxide (CO_2)

The geographical variation in CO_2 concentrations has negligible effect on the simulated BTs. However, during the ATSR missions, averaged CO_2 concentrations have increased from ~355 ppmv to ~385 ppmv. This gives a ~0.02 K decrease in the nadir 11 and 12 µm BTs, and ~0.003 K in the 3.7 µm. This is not negligible compared to the target observational stability so is accommodated in order to avoid introducing a false trend across the ATSR missions from the varying CO_2 concentration.

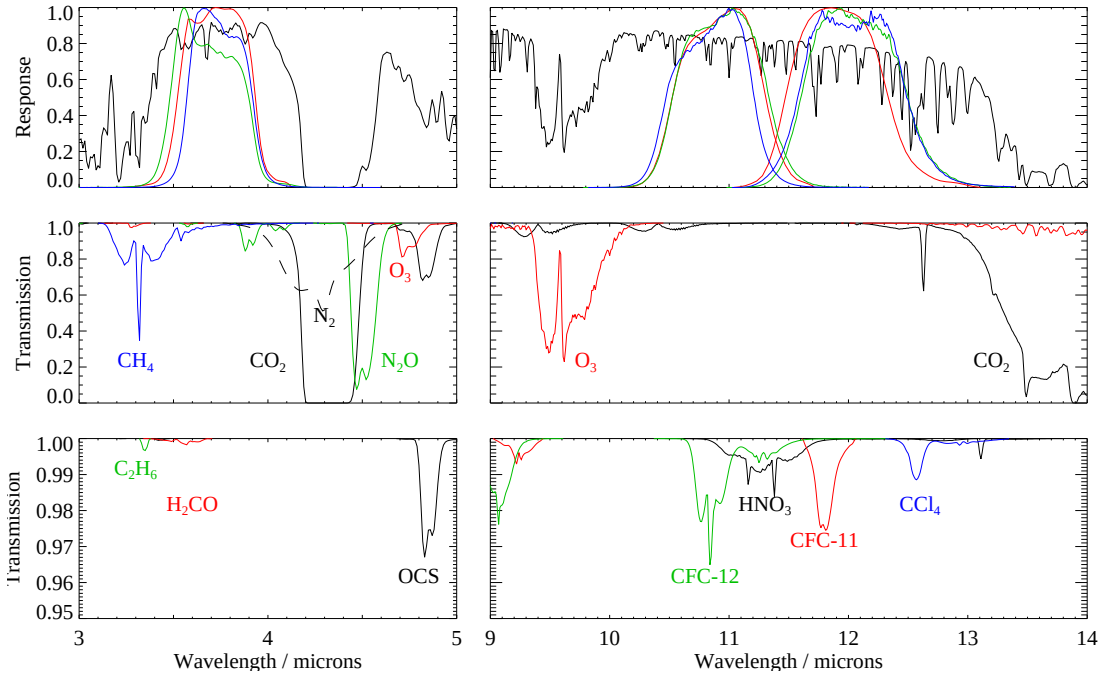

Figure 2.1: Top: ATSR spectral response functions for ATSR1 (red); ATSR2 (green); AATSR (blue) and atmospheric transmission assuming MIPAS daytime model atmosphere. Centre and bottom: atmospheric transmission for individual gases.

Gas	Difference in BT / K		
	3.7 µm	11 µm	12 µm
CO_2	−0.066	−0.200	−0.219
O_3	−0.010	−0.006	−0.029
N_2O	−0.437	−0.001	< −0.001
CH_4	−0.261	< −0.001	< −0.001
NH_3	< −0.001	−0.005	−0.001
HNO_3	0	−0.114	−0.037
OCS	< −0.001	< −0.001	−0.004
H_2CO	−0.001	0	0
N_2	−0.153	0	0
C_2H_6	< −0.001	< −0.001	-0.006
CFC-11	0	< −0.001	−0.130
CFC-12	0	−0.231	−0.006
HCFC-22	0	< −0.001	−0.019
CFC-113	0	−0.004	−0.004
CFC-114	0	−0.002	−0.002
CCl4	0	0	−0.012
HNO_4	0	< −0.001	−0.001

Table 2.1: Impact of various gases on simulated AATSR brightness temperatures. Values are changes from adding each gas, relative to a water vapour only atmosphere.

Gas	Difference in BT / K		
	3.7 µm	11 µm	12 µm
O_3	0.002	0.001	0.006
N_2O	0.036	< 0.001	< 0.001
CH_4	0.016	< 0.001	0
HNO_3	0	0.193	0.063
C_2H_6	< 0.001	< 0.001	0.006
CFC-11	0	< 0.001	0.021
CFC-12	0	0.026	0.001
HNO_4	0	< 0.001	0.002

Table 2.2: Effect of geographic variation in gas concentrations on simulated AATSR brightness temperatures. Simulations run with mid-latitude daytime temperature and water vapour profile; gas concentrations varied between 5 MIPAS model atmospheres. Values shown are maximum BT differences.

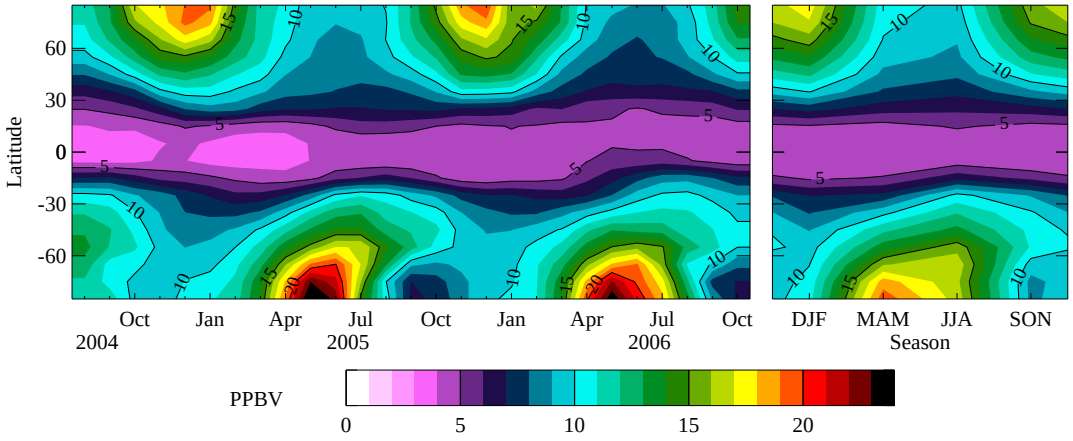

Figure 2.2: Peak stratospheric HNO_3 concentrations recorded by the MLS instrument on-board AURA. Left: time series since August 2004. Right: seasonal climatology

Records of atmospheric CO_2 from the Carbon Dioxide Information Analysis Centre (CDIAC; Keeling and Whorf, 2005) are used to scale the MIPAS model CO_2 profile to the appropriate annual concentration.

Nitric acid (HNO_3)

Nitric acid is found mainly in the stratosphere, with a maximum between 20 and 30 km. The concentrations vary from less than 5 ppbv at the equator to 15 ppbv or more in polar winter conditions. This variation in concentration is equivalent to a change of ~0.2 K in the 11 µm channel making HNO_3 the most important variable gas after water vapour.

The microwave limb sounder (MLS) instrument on the Aura satellite has been retrieving profiles of HNO_3 (Santee et al., 2007) since August 2004. We calculated monthly mean HNO_3 profiles in 10 degree latitude bins for the entire mission. The peak values of the climatological profiles are shown in Fig. 2.2 both as a monthly and quarterly climatology. The anomalously low concentrations at the south pole during October of 2005 and 2006 are due to the polar vortex.

A comparison between the averaged MLS profiles and the MIPAS model profiles showed good agreement between the two. Therefore we use the seasonal climatology of peak HNO_3 to interpolate between the MIPAS model profiles.

Nitrous oxide (N_2O) and Methane (CH_4)

Nitrous oxide and methane both have a small effect on the 3.7 µm channel which have a combined impact ~0.05 K. MLS data were again used to generate a climatology of N_2O profiles, while for CH_4 we used the Halogen Occultation Experiment (HALOE) trace gas climatology produced by Grooß and Russell (2005). There is a slight seasonal variation in gas concentrations at latitudes greater than 60 degrees, though it does not have a significant effect on simulated BTs. Therefore we simply interpolate between the MIPAS equatorial, mid-latitude and polar profiles based on latitude.

As with carbon dioxide the MIPAS profiles are also scaled to take account of global increases in trace gas concentrations. Data from CDIAC (Blake, 2005) is used for methane, and data from National Oceanic and Atmospheric Administration (NOAA) / Climate Monitoring and Diagnostics Laboratory (CMDL) for nitrous oxide. Increases in the two gases during the ATSR missions has lead to a decrease in simulated BTs ~0.02 K in the 3.7 µm channel.

CFCs

Two chlorofluorocarbons (CFC 11 and CFC 12) absorb strongly in the long wavelength channels; however latitudinal variations are equivalent to only ~0.02 K. There is also some seasonal variation in CFCs at latitudes greater than 60 degrees, particularly in the Antarctic, which is equivalent to a BT variation of ~0.005 K. However, in practice this variability is not relevant for SST estimation as such latitudes will be predominantly ice-covered during winter. After examination of Cryogenic Limb Array Etalon Spectrometer (CLAES) retrievals of CFC profiles (Roche et al., 1993) we chose to interpolate between the MIPAS equatorial and polar summer profiles based on latitude. CMDL data are used to scale the MIPAS model profiles to appropriate annual concentrations.

2.4.3 Aerosol

Marine aerosol

Marine aerosol distributions are taken from the Global Aerosol Data Set (GADS; Koepke et al., 1997) which characterises tropospheric aerosols in terms of the ten components in the OPAC dataset. Data, provided on a 5 degree grid (summer and winter separately), are surface concentration (N_0) and scale height (h) such that the particle density at height (z) is given by

$$N(z) = N_0 e^{-z/h} \tag{2.4}$$

The components present are water-soluble and sea salt components for clean maritime conditions; while coastal, polluted areas will also include soot.

Stratospheric aerosol

Due to the eruptions of Mount Pinatubo in June 1991, and Mount Hudson in August of that year, there have been significant changes in the nature of the stratospheric aerosol throughout the ATSR missions. The eruption caused ~50× increase in aerosol optical depths which then decayed until reaching background conditions around 1997. However, in addition to simply increasing the optical depth, the microphysical properties were also affected. The injection of massive amounts of sulphur dioxide gas affected the formation of the resulting sulphuric acid aerosol (H_2SO_4-H_2O) leading to an increase in the average aerosol particle size, which in turn will affect the optical properties of the aerosols.

Deshler et al. (2003) have published in situ size distribution measurements of stratospheric aerosol made at Laramie, Wyoming during the last 30 years. These data provide vertical profiles of both number density and the size distribution fitted using a bimodal lognormal function. Each mode is described by: the number density (particles / cm^3), median particle radius (μm), and distribution width (dimensionless). The primary mode represents the more numerous small particles. However, the larger particles of the secondary mode have a greater impact on infrared scattering properties.

Aerosol profiles were smoothed with a median filter and the evolution of number density and particle radius at various heights is plotted in Fig. 2.3. The number density of the primary mode shows relatively modest change associated with the Mount Pinatubo eruption. However, the radius does increase briefly after the eruption, doubling to ~0.15 μm for a year or so. The effects on the secondary mode are more dramatic. The secondary number density increases by one or two orders of magnitude at all altitudes, while the particle radius at lower altitudes increases to just under 0.6 μm. After the eruption the effects on the secondary mode persist for several years, not decaying to background levels until 1997/1998.

Figure 2.4 shows profiles of the bimodal number density and partical radius for background conditions (1997 to 2004) and elevated aerosol conditions after Pinatubo (secondary mode only). For purposes of RT modelling the primary mode number density is fitted using Eq. 2.4 with N_0 = 21460.9 cm^{-3} and h = 2.85921 km for altitudes 12–24.55 km; above this within the range 24.55–32 km the number density is held constant at $N(z) = 4$ cm^{-3}. Outside this altitude range the number density is set to zero. Particle radius is low ~0.025 μm, but with a band at ~25 km altitude of larger particles. For a year or so after the eruption, primary particle radii increased with the height of maximum radius being slightly lower while particle number density is relatively unchanged.

For the secondary mode the concentration also varies from year to year, however it was sufficient to use one number density profile and scale it depending on the year. The number density profile, fitted to 1990 data, is assumed to be consant $N(z) = 0.173$ cm^{-3} for 12–24 km and exponentially

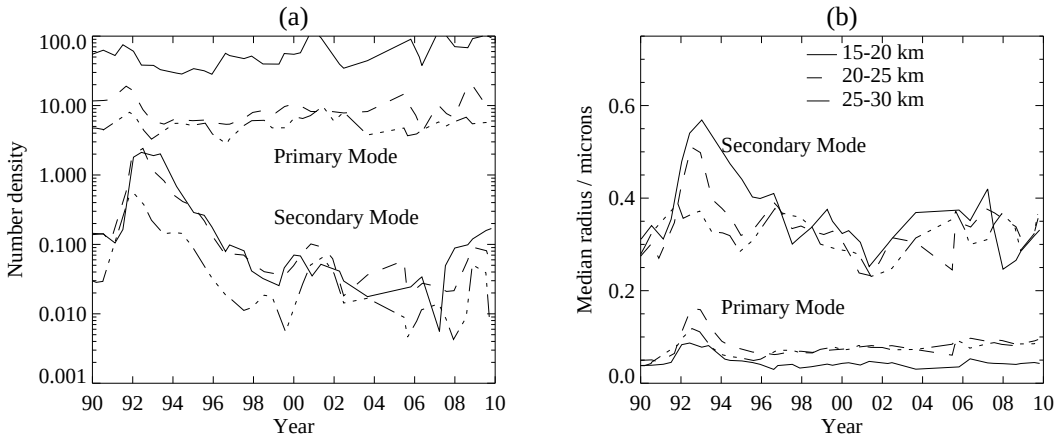

Figure 2.3: Evolution of stratospheric aerosol number density and particle radius after Pinatubo eruption (June 1991).

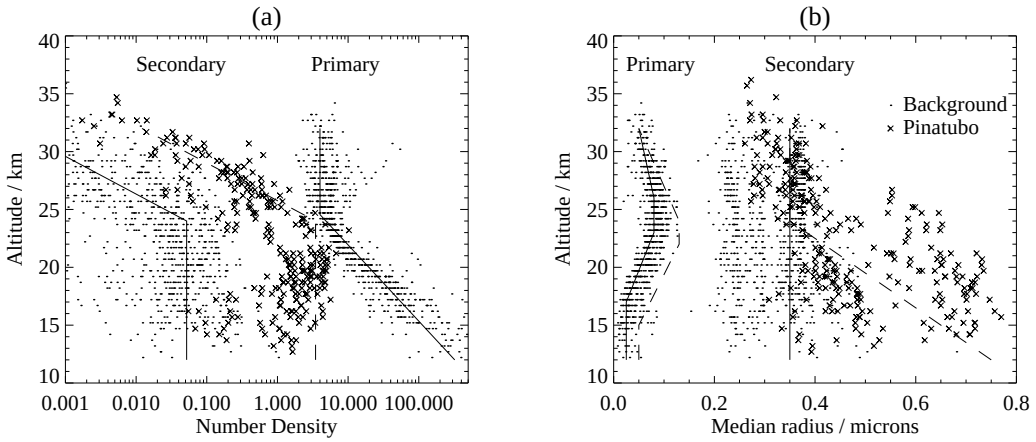

Figure 2.4: Stratospheric aerosol number density (a) and particle radius (b) for volcanic/Pinatubo (1992) and background (1997–2004) conditions. Points indicate background measurements and solid line is model fit. Crosses indicate measurements of secondary mode during 1992 and dashed line is model fit.

Year	1991	1992	1993	1994	1995	1996	1997+
Scale	1.0	20.0	9.0	4.0	1.5	0.5	0.3

Table 2.3: Time dependent scaling applied to stratospheric aerosol secondary mode number densities.

decreasing up to 32 km ($N_0 = 4.04370 \times 10^6$ cm^{-3} and $h = 1.41467$ km). The whole profile is then scaled by the year-dependent factors shown in Table 2.3. Under background conditions the secondary mode is considered to have a constant radius of 0.35 μm at all altitudes. While volcanic aerosol is still present particle radii are larger at lower altitudes, this is modelled as varying linearly from 0.35 μm at 24 km to 0.75 μm at 12 km for 1992/93, and 0.55 μm for 1994/95.

2.5 Simulation results

The set of atmospheric states from section 2.4 were processed with the radiative transfer model of section 2.3 to produce top-of-atmosphere brightness temperatures and aerosol ΔBTs for the various ATSR instruments at a range of viewing angles and times (to represent changing trace gas concentrations). Viewing angles were chosen to correspond to path lengths, $s = \sec\theta$, of 1.00, 1.02, 1.04, 1.06, 1.08 for the nadir view; and 1.60, 1.64, 1.68, 1.72, 1.76, 1.80 for the forward view. Trace gas concentrations used were appropriate for four year intervals corresponding to: 1991, 1995, 1999 etc. The following sections compare the stratospheric aerosol ΔBTs with those calculated by Merchant et al. (1999) and the clear-sky BTs with empirical observations.

2.5.1 Stratospheric aerosol

To date there have been no volcanic eruptions large enough to affect stratospheric aerosol during the life of the AATSR instrument and the stratospheric perturbation caused by the Pinatubo eruption had largely decayed by the time ATSR2 had begun operations; however, SST retrieval coefficients for ATSR2 and AATSR have been made robust to volcanic aerosol. This robustness is achieved though choosing the retrieval coefficients for Eq. 2.1 such that $\mathbf{a}^{\mathrm{T}}\mathbf{k} = 0$ where $\mathbf{k}$ is a vector representing the mode of variability of the aerosol – i.e. the average ΔBT for each channel. As the stratospheric aerosol calculations of Merchant et al. (1999) were effective in reducing aerosol related biases in ATSR1 data, we compare our simulated ΔBTs against theirs.

Merchant et al. (1999) considered three stratospheric aerosol types when examining the effects of Pinatubo aerosol on ATSR1 retrievals. These were the three available in the radiative transfer code MODTRAN3: “fresh,” “aged” and “background” corresponding to the aerosol present immediately after the eruption, about a year after, and many years after respectively. They neglected the fresh aerosol type as ATSR1 began routine operations a few months after the eruption and the radiative impacts were found to be very similar to the aged type.

For comparison we calculate ATSR1 aerosol ΔBTs for background (1991, pre-Pinatubo) and maximally elevated aerosol (1992) conditions. These two years represent the most widely different aerosol

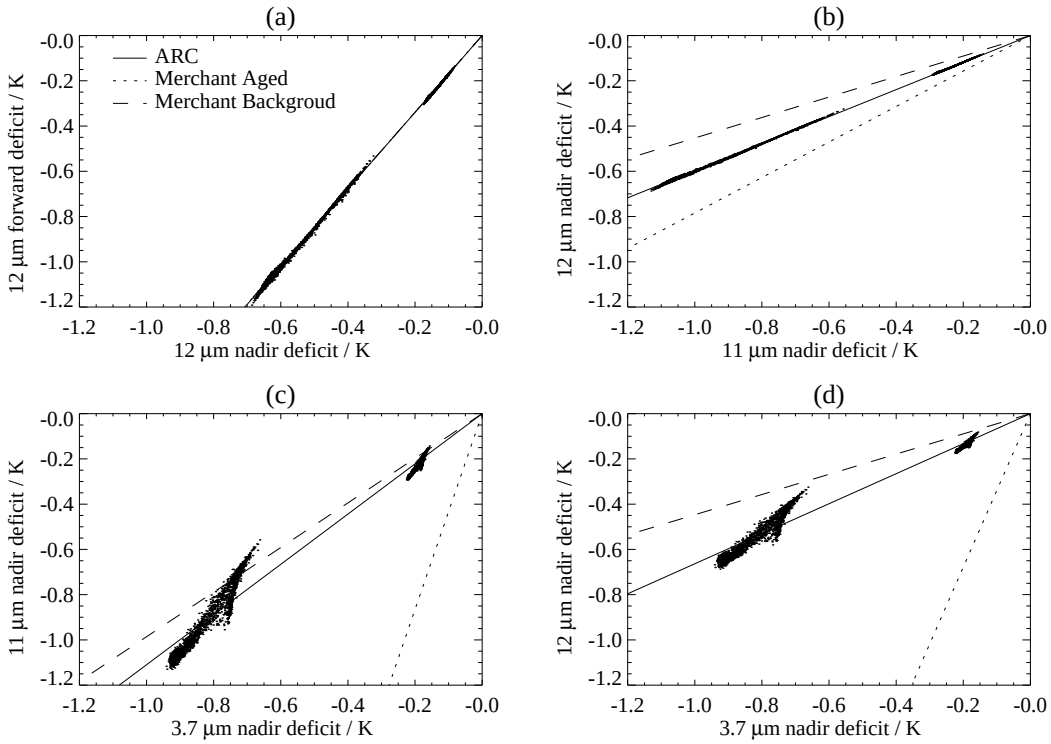

Figure 2.5: Simulated change in brightness temperature due to the presence of stratospheric aerosol for volcanic and background aerosol distributions. Background simulations (small cluster) have been scaled by ×10. Each cluster has fixed aerosol, but varying meteorological state. Solid line shows average mode of variablity for simulations. Also shown are aged (dotted line) and background (dashed line) aerosol modes from in previous work (Merchant et al., 1999).

properties — 1997 onwards had slightly lower aerosol densities than 1991, but the same micro-physical properties. As such we would expect the "aged" aerosol of Merchant et al. (1999) to most closely match our elevated aerosol (1992).

Figure 2.5 shows the aerosol ΔBTs from the new simulations compared with the aged volcanic and background aerosol modes of Merchant et al. (1999). The background aerosol ΔBTs (~0.02 K) are scaled by ×10 for readability. The old "aged" and "background" modes differ significantly in the inter-channel ratios, especially 3.7 µm against 11 and 12 µm. However, the new simulations vary predominantly in magnitude, not relative impact on channel, between background and elevated conditions despite the different micro-physical properties for the two. Figure 2.6 shows the mode of variation (**k**) calculated for the new simulations along with those from Merchant et al. (1999). Each mode has been normalised such that the magnitude of the vector **k** is one.

As stratospheric aerosol is optically thin, there is a near-linear relationship between optical depth and transmission, consequently the relation between nadir and forward view ΔBTs is a function of the viewing geometry rather than the aerosol optical properties. This is shown for the 12 µm channel in Fig. 2.5a where the forward:nadir ratio is near-identical (1.70) for all new simulations and the previous aerosol simulations from Merchant et al. (1999). The deficit ratio changes with the viewing geometry becoming 1.53 at the edge of swath; other channels have equally linear relations, but slightly different

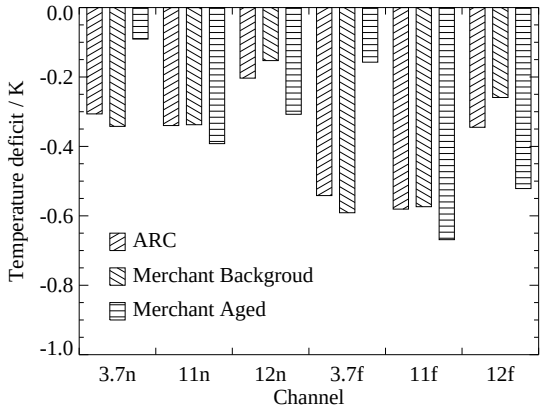

Figure 2.6: Modes of variation introduced by stratospheric aerosol. Temperature deficits are normalised to have magnitude of 1 K.

ratios: 1.77 and 1.59 for centre and edge of swath in the 3.7 µm channel. For comparison the ratio of path length at centre and edge of swath is 1.76 and 1.58.

Figure 2.5b shows the 11 and 12 µm ΔBTs are also highly correlated, the slope here is independent of viewing angle. The previous aged and background aerosol modes had moderately different impacts on the 11 and 12 µm channels, with the new simulations lying approximately half way between the two.

Figure 2.5c and 2.5d compare the 11 and 12 µm channels to the 3.7 µm channel, again the variation of optical properties has little effect beyond changing the amplitude of the ΔBTs. However, the previous aged and background modes were significantly different, with the background mode being closest to the new simulations based on measured aerosol size distributions.

While the change in aerosol micro-physical properties does not affect the BT deficit ratios, varying the meteorological state does. This can be seen in Fig. 2.5c and 2.5d where the cluster of simulations spreads out from the "best-fit" line. This is primarily due to the greater non-linearity of the Planck function in the near infrared compared to thermal infrared, combined with the range of surface temperatures modelled (hence varying upwelling radiances). For low surface temperature, corresponding to the smaller ΔBTs, thermal emission from the aerosol itself can reduce the BT11/12 deficit; in the 3.7 µm channel, thermal emission from the aerosol is negligible due to the non-linearity of the Planck function. Effectively, BT11/12 is more sensitive to the temperature of the aerosol than BT3.7 is.

As the evolution of the aerosol properties only affects the magnitude of the BT deficit, and not the ratio between channels, we only require one aerosol mode to represent the observed Pinatubo aerosol. Unlike the previous "aged" and "background" optical properties, the new aerosol simulations are based on actual measurements of stratospheric aerosol particles (Deshler et al., 2003).

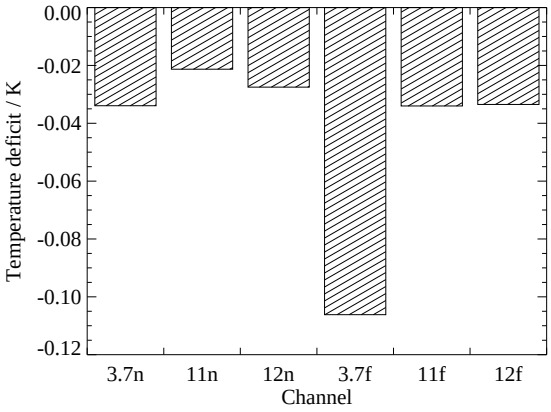

Figure 2.7: Average simulated BT deficit due to marine aerosol

2.5.2 Marine aerosol

Unlike stratospheric aerosol where there is a clearly defined mode of variability and the ΔBT impacts are (during volcanic periods) large, the effects of marine aerosol cannot be defined using a single mode and the ΔBTs are typically < 0.1 K. Therefore the effects of marine aerosol are included in the clear-sky BTs that are compared against observations in the following section. Here, we isolate the simulated effects of marine aerosol on BTs.

Figure 2.7 shows the global average marine ΔBTs for each channel, the effect in the longwave is a ~0.03 K decrease in BT; however, this is quite variable and it is possible for aerosol to actually increase the simulated BT under the correct meteorological conditions. As with the stratospheric aerosol, the 11 and 12 μm ΔBTs are highly correlated as are the forward and nadir ΔBTs. The forward:nadir ratio for the 11 μm channel is ~1.6, similar to the stratospheric aerosol case. For the 12 μm channel it is significantly lower at ~1.2, this is due to the higher water vapour absorption in this channel from tropospheric water vapour. The effects in the 3.7 μm channel, however, are very different; firstly the 3.7 μm ΔBTs are not strongly correlated to the 11 and 12 μm, secondly the forward:nadir ratio is much larger ~3.1. This is due to scattering being the dominant process at 3.7 μm while at longer wavelengths it is negligible. For the conditions simulated, the marine aerosol optical depths are low and scattering in the 3.7 μm channel is predominantly into the forward direction. Hence most of the scattered radiance reaching the sensor in the nadir-view originated at the surface, whereas in the forward-view more of the scattered radiance originated as down-welling atmospheric emission[1]. The lower water vapour absorption in the 3.7 μm channel means the aerosol effects are more easily detectable than in the 11 and 12 μm channels. However, water vapour alone cannot account for the large forward:nadir ratio which would be comparable to the path length ratio (1.76) if scattering was not significant.

Figure 2.8 shows the spatial variability of the nadir 3.7 μm and 11 μm deficits; the spatial variability

[1] Down-welling radiance emitted by the atmosphere is much lower than the upwelling radiance emitted by the surface as the atmosphere is colder than the surface and is partially transparent.

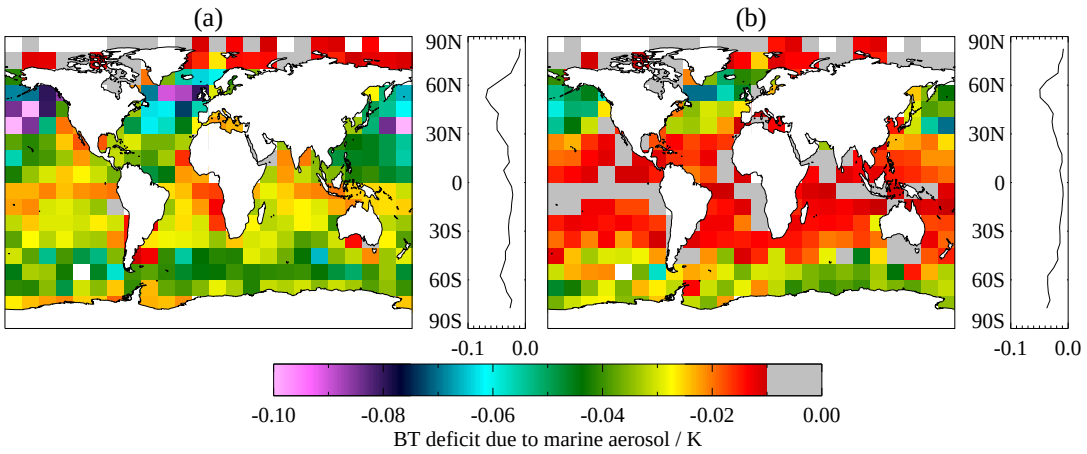

Figure 2.8: Geographical variation of BT deficit due to marine aerosol for (a) 3.7 µm and (b) 11 µm channels.

of the 12 µm channel, not shown, is very similar to the 11 µm, and the forward view deficits follow the same distribution as the nadir but with higher magnitude. The greatest aerosol effects are predicted in the northern hemisphere – particularly mid-latitudes over the Atlantic and Pacific Oceans – reflecting the higher aerosol concentrations compared with the southern hemisphere

2.5.3 Emissivity

To quantify the impact of the new emissivity model, top-of-atmosphere BTs were calculated using both the new emissivity model and that of Merchant et al. (1999). The RFM transmittance model (subsection 2.3.1) along with the updated profile set (section 2.4) was used for both sets of calculations. Figure 2.9 shows the change in BT (new model - previous) against SST for AATSR at nadir. The box and whisker plots show that the distribution of BT deficits in the colder SST bins is asymmetric, particularly for the 12 µm channel; this is due to outliers where the absorption from water vapour has masked the deficit. At temperatures above 300 K the differences between the two emissivity models is negligible with mean differences $\lesssim 0.01$. However, at lower temperatures the new refractive index data leads to lower emissivities and lower BTs for all channels leading to a maximum difference of ~0.3 K in the 12 µm channel consistent with results of Newman et al. (2005).

The use of the old surface emissivity when generating the SST retrieval coefficients leads to an over estimation of the retrieved SST at low temperatures. This occurs as the retrieval algorithms are sensitive to the $11-12\,\mu$m difference which has been increased with the new emissivity. The BT deficits shown in Fig. 2.9 imply that the operational ATSR coefficients are over estimating the SST at low (< 275 K) temperature by approximately 0.42, 0.11, 0.14, 0.05 K for the N2, N3, D2, and D3 retrievals respectively. In practice, however, operational ATSR SST retrievals appear to have cold biases in polar regions. This is due to other errors in the original RT simulations. For instance, the regional effects of trace gases primarily affect the 11 µm channel and counteract the surface emissivity errors to some extent.

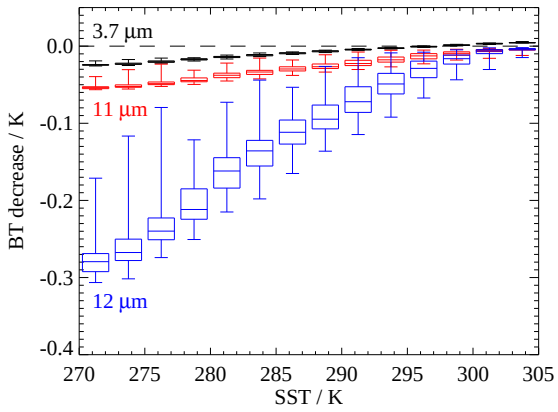

Figure 2.9: Change in top-of-atmosphere BT due to new emissivity model. Box and whisker plots show minimum, lower quartile, median, upper quartile, and maximum in each bin of SST.

2.5.4 Clear-sky BTs

As the simulated BTs are to be used for generating SST retrieval coefficients it is important that they accurately represent global, clear-sky, observations and associated SSTs. Any bias in the simulated BTs will lead to biases in the retrieved SST (Merchant and Le Borgne, 2004). In this case the simulation biases include not only the errors directly attributable to the radiative transfer, but also those associated with the input atmospheric profiles – for example systematic biases in the temperature profile, or an undersampling of certain atmosphere conditions.

In this sub-section we evaluate the suitability of the simulated BTs by comparing them with empirical distributions of clear-sky BT observations. However, direct comparison of the simulated and observed BT distributions is not particularly informative as they are dominated by variations in surface temperature while atmospheric effects and subtle differences in the forward modelling cannot be distinguished. We can eliminate the majority of the variability from SST by considering the BT differences between channels (the majority of SST retrieval formulations are explicitly written in terms of BT differences rather than the more general form shown in Eq. 2.1). The approach we take is to produce BT difference plots – i.e. 3.7 μm – 11 μm plotted against the nadir-forward difference in the 3.7 µm channel. As, for example, errors in the forward model will affect the various channels and view in different ways, any offset between the simulations and observations in BT difference plots is an indication of a weakness in some aspect of the simulations.

The clear-sky observations were extracted from the (A)ATSR Multimission Archive during a trial run of the ATSR Reprocessing system (2002–2008), along with co-located NWP model data (ERA-40 up to September 2002, and ECMWF operational data afterwards). Observations were filtered to select only those which were night-time, clear-sky, and centre-swath. For clear-sky we required all 25 pixels in a 5 × 5 box centred on the target to be classified as clear by the cloudmask. Centre-swath includes pixels within 140 km of the sub-satellite track (i.e. nadir zenith angle less than 11.37°).

The distribution of ATSR2 observations are shown in Fig. 2.10, with sub plots a–c showing the

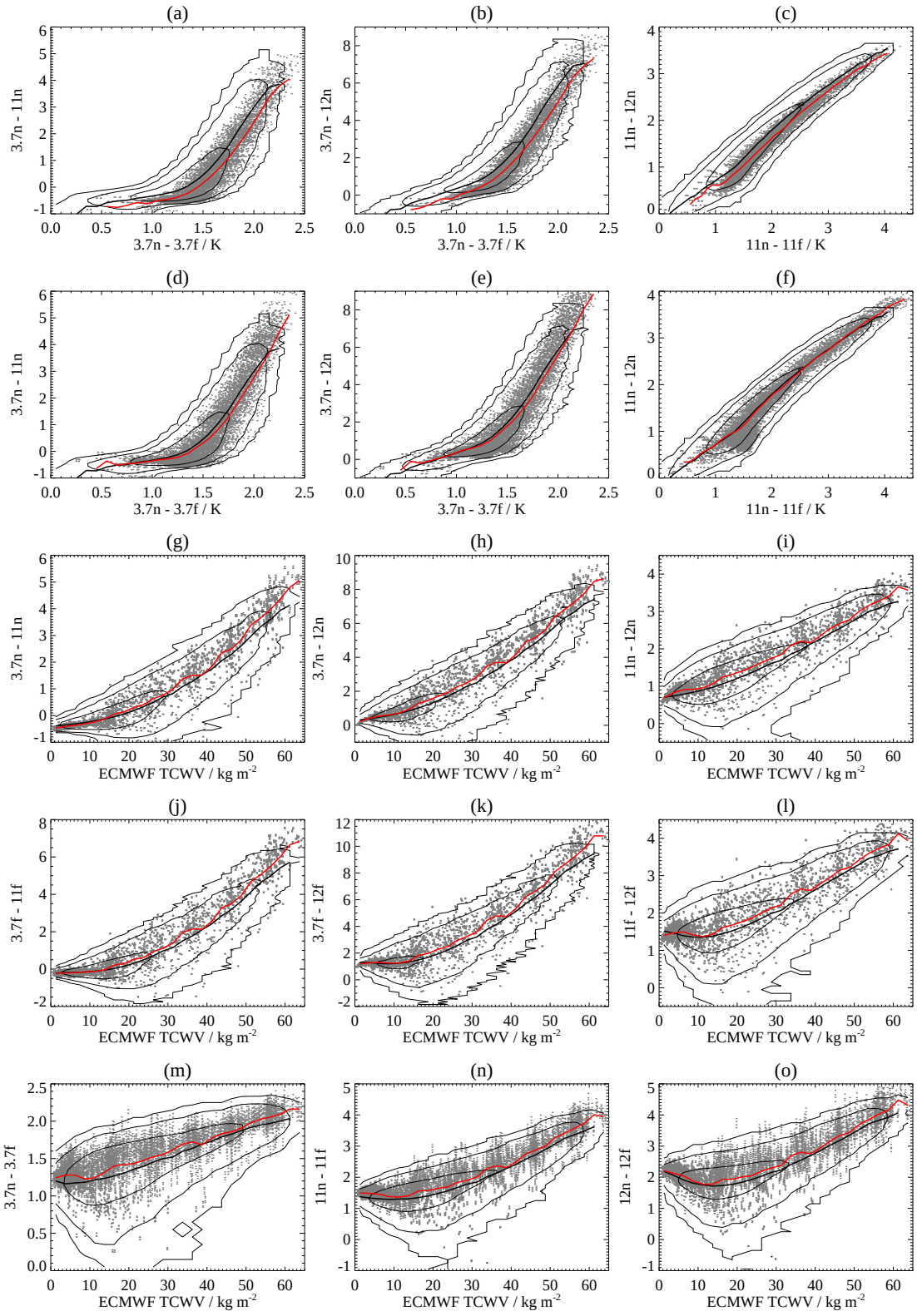

Figure 2.10: ATSR2 BT difference plots comparing simulations and observations. Contours - distribution of observations; points - simulated BTs; thick solid line - 'best fit' through observations; dashed (red in online version) line - 'best fit' through simulations. (a,b,c) Simulations used in previous work (Merchant et al., 1999). (d,e,f) New RFM simulations. (g,h,i) RFM nadir view combinations. (j,k,l) RFM forward view combinations. (m,n,o) RFM nadir-forward differences.

simulations used for previous work (Merchant et al., 1999, hereafter, M99), and sub plots d–f show the new RFM simulations. In the M99 plots there is a clear offset between the centres of the observation distribution and the simulations, while the RFM simulations significantly improve the agreement. In the case of the 11 and 12 µm differences, compare Fig. 2.10c and Fig. 2.10f, there is an offset ~0.2 K between the observed 11-12 μm difference and the M99 simulated difference. The new RFM simulations are in very close agreement with the empirical distribution with the double difference now <0.1 K. When comparing the 3.7 µm channel with the split-window (Fig. 2.10a,b and Fig. 2.10d,e) the M99 simulations show offsets of ~0.2 K and ~0.4 K, for the 3.7-11 and 3.7-12 double differences respectively, for low nadir-forward differences (<0.1 K). The RFM simulations show an improvement in the agreement: offset <0.1 K for low nadir-forward differences increasing to ~0.2 K for the largest nadir-forward difference. The lowest nadir-forward differences, i.e. $3.7n - 3.7f < 1.0$, were poorly represented in the M99 simulations, these correspond to polar conditions with large (positive) air-sea temperature differences, but are present in the new profile set.

The remaining plots in Fig. 2.10 show the various BT differences against corresponding ECMWF total column water vapour (TCWV) for the new simulations: nadir view combinations in sub plots g,h,i; forward view combinations in j,k,l; and view differences in m,n,o. For the single view combinations there is excellent agreement between simulations and observations with double differences $\lesssim$0.1 K for drier conditions (TCWV < 20 kg m^{-2}).For higher water vapour loadings there is a slight divergence between simulation and observation up to ~0.1 K for the $11n - 12n$ and ~0.5 K for the $3.7f - 12f$ double difference. This is due to the presence of simulations with very large $3.7f - 12f$ differences, these represent conditions with very low 12 µm transmittances which have most likely been classified as cloud in the observational data. From the view difference plots in Fig. 2.10m,n,o it appears the simulations have slightly higher nadir-forward differences than the observations by 0.1–0.2 K on average. However, the simulations do cover the full range of nadir-forward differences seen in observations. When considering all the plots against TCWV together it is evident that the simulations are over-estimating the channel and view differences at higher water vapour loadings for all the combinations considered and could be brought into agreement by scaling the TCWV axis for one of the simulations or observations. This would imply that either the radiative transfer simulations are over estimating the impact of water vapour or the ECMWF TCWV co-located with the observations is an over estimation of the true TCWV.

Figure 2.11 is the AATSR equivalent to Fig. 2.10 and the same general behaviour is seen. However, in the M99 $11n - 12n$ plot (Fig. 2.11c) there is now a large ~0.4 K offset between observation and simulation, this is reduced to ~0.2 K when using the RFM simulations shown in Fig. 2.11f. However, there is a known ~0.2 K cold bias in the AATSR 12 µm channel (G Corlett, personal communication). This appears to be an instrument calibration issue – hypotheses include out-of-band detector response and shifted spectral response function – and work is under-way to identify and characterise the problem (Dave Smith, personal communication). Figure 2.12 shows the same channel combinations as Fig. 2.11f with the recommended correction applied to the 12 µm channel, this has eliminated

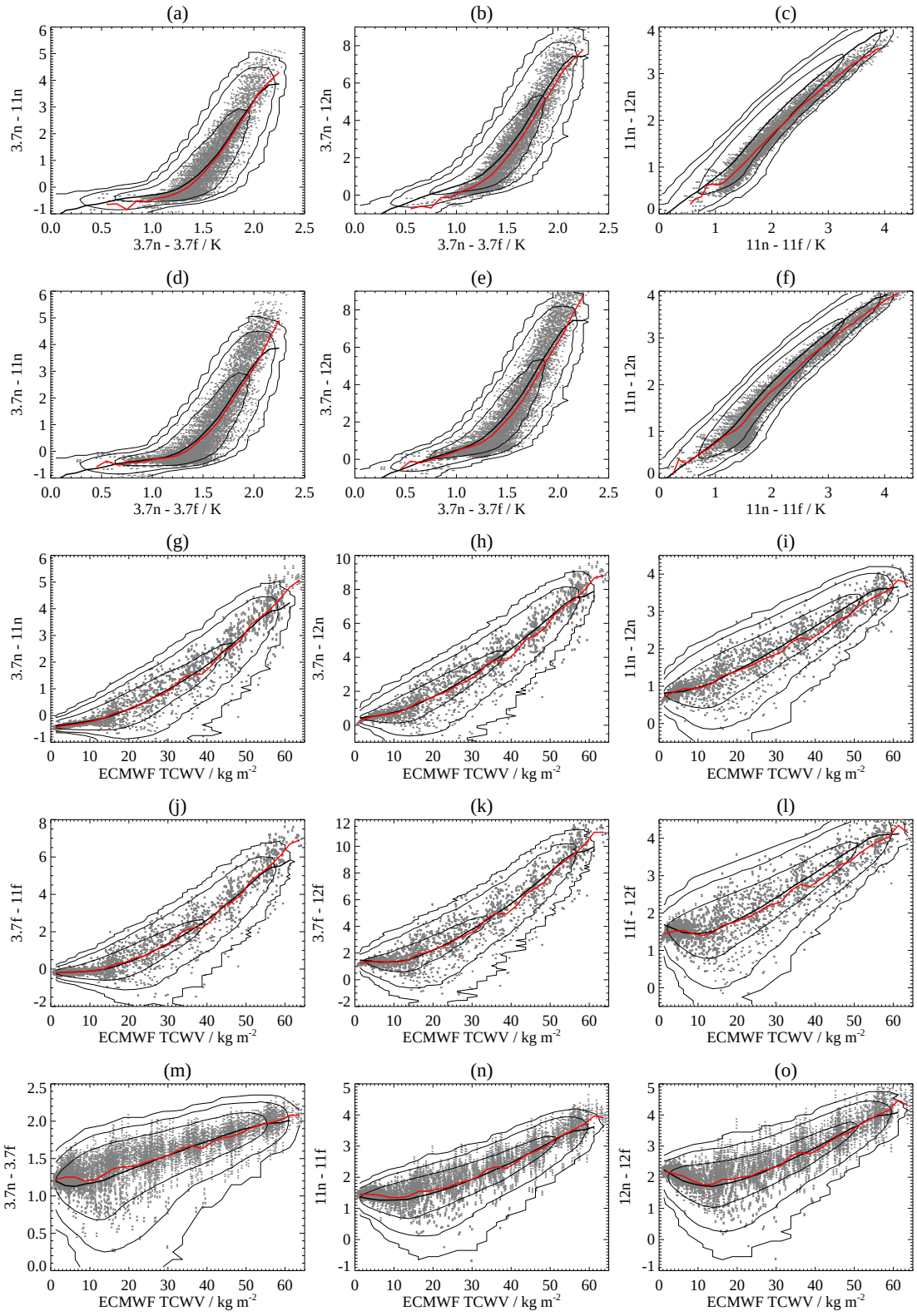

Figure 2.11: AATSR BT difference plots comparing simulations and observations. Contours - distribution of observations; points - simulated BTs; thick solid line - 'best fit' through observations; dashed (red in online version) line - 'best fit' through simulations. (a,b,c) Simulations used in previous work (Merchant et al., 1999). (d,e,f) New RFM simulations. (g,h,i) RFM nadir view combinations. (j,k,l) RFM forward view combinations. (m,n,o) RFM nadir-forward differences.

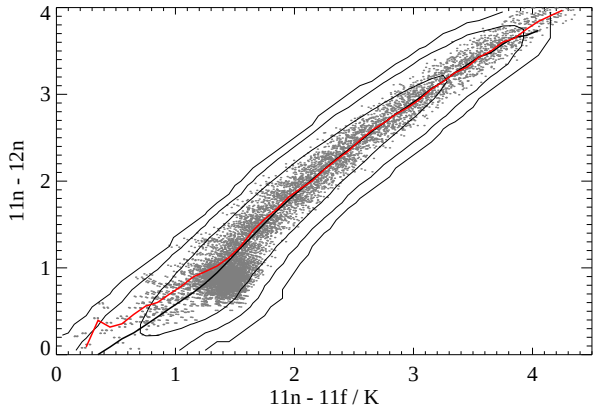

Figure 2.12: AATSR BT difference plot including 0.2 K correction to 12 µm channel. RFM simulations show. Key as Fig. 2.11

the offset for the majority of cases ($11n - 11f > 1.5$ K) but has over-corrected in drier conditions ($11n - 11f < 1.5$ K). In the plots of single-view channel combinations against TCWV Fig. 2.11g–l there is generally better agreement with observations than seen in the ATSR2 results. For the $3.7n - 11n$ (Fig. 2.11g) and $3.7f - 11f$ (Fig. 2.11j) the double differences remain $\lesssim$0.1 K for the full range of TCWV. While channel combinations involving the 12 µm channel have the ~0.2 K offset in wetter atmospheric conditions.

Compared to the old M99 simulations the updated RFM simulations show better agreement with observations in BT difference plots. This is evident in both the relative offset of simulated distribution to observed and in the range of conditions covered. These comparisons highlight the 0.2 K bias previously found to affect the AATSR 12 µm channel. The simple constant correction seems effective for the majority of conditions (Fig. 2.12). Aside from this 12 µm issue, AATSR simulations-observations differences appear to have a lower dependence on TCWV than do ATSR2 simulations. Assuming that the simulation quality is equivalent for both sensors and there are no other calibration biases affecting ATSR2 or AATSR, this could imply some systematic difference in the TCWV values used to generate the observation distributions. As the observations are co-located with NWP TCWV data, a systematic bias in these TCWV values would result in the observation contours in Fig. 2.10 and Fig. 2.11g-o being shifted horizontally — i.e. the relation between TCWV and BT differences would change for the observations, but not the simulations. As TCWV values are taken from ERA-40 up to September 2002 and ECMWF operational afterwards, the ATSR2 observations mainly use ERA-40 while the AATSR observations use ECMWF operational. Reducing the ERA-40 TCWV values by ~5% (i.e., rescaling by a factor of 0.95) has the effect of bringing the ATSR2 simulations and observations into a level of agreement comparable to that seen for AATSR. However, we know of no independent justification for such a scaling.

2.6 Summary and Conclusions

A new dataset of simulated top-of-atmosphere BTs and associated SSTs have been created for use in generating SST retrieval coefficients within the ATSR Reprocessing for Climate project. The approach taken includes the use of a full line-by-line radiative transfer model along with recent spectroscopic data, and the DISORT computer code for scattering effects.

An updated set of atmospheric profiles based on the dataset from Chevallier (2002), representative of the temperature and water vapour simulated in the ERA-40 analysis, provides the primary meteorological dataset. The previous approach of filtering profiles based on surface temperature to exclude cases where $T_{\text{surf}} < 271.35\,\text{K}$ lead to an under-representation of high-latitude cases as the NWP surface temperature is a grid cell average and therefore includes both sea and sea-ice surface temperatures. By generating sea-surface temperatures using a simple empirical model based on the SST to ice-coverage relation from HadISST we can retain high-latitude profiles with larger air-sea temperature differences.

All atmospheric gases where spatial/temporal variation in gas concentrations has an impact on simulated top-of-atmosphere BTs greater than 0.01 K are added to the profile dataset; atmosphere gases which have an impact of simulated BT greater than 0.001 K are included as fixed gases. Compared to the earlier radiative transfer simulation of Závody et al. (1995) and Merchant et al. (1999) we add the following fixed gases: H_2CO, C_2H_6, HCFC-22, CFC-113, CFC-114, CCl4, and HNO_4. Taken together this accounts for a ~0.04 K effect on the 12 µm channel. In addition, geographical variations in some trace gases are now included: N_2O, CH_4, HNO_3, and CFC-11 and CFC-12. These cause zonal and, in the case of HNO_3, seasonal variations of order $0.05, 0.2, 0.1\,\text{K}$ in the $3.7, 11, 12\,\mu$m channels respectively. Secular trends in atmospheric concentrations of CO_2, N_2O, CH_4, and CFC-11 and CFC-12 are included as these lead to drifts in simulated BTs ~0.02 K through the course of the ATSR missions.

While the eruption of Mount Pinatubo had a significant impact on the microphysical and optical properties of stratospheric aerosol (Massie et al., 1996), the relative BT impacts between ATSR channels is now found to be unchanged, meaning that a single mode of BT variability is sufficient to account for the aerosol effects, rather than two as used in Merchant et al. (1999). Effectively, the change in aerosol optical properties are indistinguishable from changes in optical depth using just the ATSR infrared channels. When considering just the 11 and 12 µm channels, the new aerosol mode is, approximately, the average of the "aged" and "background" modes used by Merchant et al. (1999), while the impact on the 3.7 µm channel is closer to the "aged" mode. Identifying a single aerosol mode to represent the stratospheric variation has two advantages. Firstly, it becomes possible, in principle, to generate single-view night-time SST retrieval coefficients robust to a single mode. Secondly, if an independent estimate of the aerosol index, or infrared optical depth, is available then it becomes possible to estimate the bias as $\mathbf{a}^{\text{T}}\mathbf{k}$ and correct retrievals from sensors such as AVHRR. These possibilities will be explored in future work.

Initial comparison with observed ATSR2 and AATSR BTs show better agreement with the new

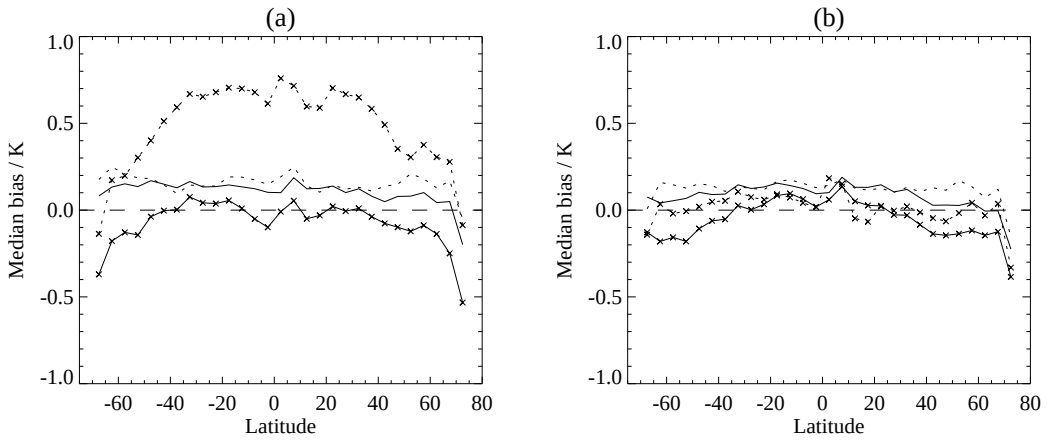

Figure 2.13: Median retrieval bias (satllite-in situ) for ATSR2 using (a) operational coefficients, and (b) coefficients derived from updated RT modelling. Dashed with symbol - N2; dashed - N3; solid with symbol - D2; solid - D3.

simulations than those used in previous ATSR work. In double difference plots, offsets between observed and simulated distributions been have decreased from ~0.2-0.5 K to ~0.1 K. Simulations of accuracy <0.1 K are necessary in order to generate SST retrieval coefficients accurate to ~0.1 K.

A companion paper (Embury and Merchant, 2012) describes a new retrieval scheme based on the radiative transfer results presented in this paper. But it is possible to show the improvement from the new simulations by generating retrieval coefficients using the same methodology as the operational coefficients. Figure 2.13 shows retrieval bias with respect to in situ drifting buoys for the operational ATSR2 retrieval coefficients and a set of coefficients based on our new radiative transfer results. For the operational coefficients the inter-algorithm biases (i.e. between different channel combinations) are up to ~0.6 K similar to the double difference offsets observed for the old RT simulations. The largest biases, relative to in situ, affect the nadir-view two-channel (N2) retrieval while the dual-view three-channel (D3) shows the lowest bias and variation with latitude. Using the updated coefficients the inter-algorithm biases are reduced to ~0.1–0.2 K and the latitudinal variability affecting the two-channel N2 and D2 retrievals is reduced.

Using the improved SST retrieval algorithm, described in the companion paper (Embury and Merchant, 2012), the inter-algorithm biases are improved to order ~0.1 K, verifying the accuracy of the simulations. A further paper (Embury et al., 2012b) presents an initial comparison of the satellite retrieved SST against in situ measurements. This work will form the basis of the SST retrieval in the ARC project. Further work will include assessment of the long term stability in order to ensure measured trends are accurate to 0.05 K decade^{-1}, and analyses of the overlap periods between sensors to homogenise the record and remove residual errors arising from remaining forward modelling uncertainties.

2.7 Acknowledgements

The ATSR Reprocessing for Climate project is jointly funded by the Natural Environment Research Council, the Ministry of Defence, and the Department of Energy and Climate Change (previously Defra). Archived ATSR data and computer facilities are provided by the NERC Earth Observation Data Centre (http://www.neodc.rl.ac.uk).

This work makes use of the RFM radiative transfer model developed by Anu Dudhia (University of Oxford) in combination with the TES line parameter database (AER Inc.); NWP data from the European Centre for Medium-Range Weather Forecasts; MIPAS model atmospheres compiled by John Remedios (University of Leicester) and trace gas trends from the Carbon Dioxide Information Analysis Center, and the NOAA/Earth System Research Laboratory (formerly CMDL); and stratospheric aerosol size distribution measurements supplied by Terry Deshler (University of Wyoming).

Chapter 3

A New Retrieval Scheme

Owen Embury, Christopher J. Merchant

Abstract

We present a new coefficient-based retrieval scheme for estimation of sea surface temperature (SST) from the Along Track Scanning Radiometer (ATSR) instruments. The new coefficients are banded by total column water vapour (TCWV), obtained from numerical weather prediction analyses. TCWV banding reduces simulated regional retrieval biases to <0.1 K compared to biases ~0.2 K for global coefficients. Further, detailed treatment of the instrumental viewing geometry reduces simulated view-angle related biases from ~0.1 K down to <0.005 K for dual-view retrievals using channels at 11 and 12 μm. A novel analysis of trade-offs related to the assumed noise level when defining coefficients is undertaken, and we conclude that adding a small nominal level of noise (0.01 K) is optimal for our purposes.

When applied to ATSR observations, some inter-algorithm biases appear as TCWV-related differences in SSTs estimated from different channel combinations. The final step in coefficient determination is to adjust the offset coefficient in each TCWV band to match results from a reference algorithm. This reference uses the dual-view observations of 3.7 and 11 μm. The adjustment is independent of in situ measurements, preserving independence of the retrievals. The choice of reference is partly motivated by uncertainty in the calibration of the 12 μm of Advanced ATSR.

Lastly, we model the sensitivities of the new retrievals to changes to TCWV and changes in true SST, confirming that dual-view SSTs are most appropriate for climatological applications.

3.1 Introduction

This paper presents the definition of a new retrieval scheme for the estimation of sea surface temperature (SST) from the Along Track Scanning Radiometers (ATSRs). The scheme is directly applicable to

ATSR-2 and Advanced ATSR (AATSR). For ATSR-1 there are additional factors that necessitate separate treatment in a later paper.

The retrieval scheme is based on defining coefficients for a weighted combination of ATSR observations to form an estimator for the sea surface temperature. This general approach is computationally efficient, effective and well established. Many accounts have been published defining coefficients for various sensors, using various "algorithms", dedicated to particular regions, etc, a sample of which are cited below. The remainder of this introductory section gives a thematic overview placing the present work in this context. Thereafter, the paper progresses as follows. Section 3.2 sets out the particularities of SST retrieval from ATSRs and outlines the current operational retrieval scheme. Section 3.3 develops the new retrieval scheme using simulations. Section 3.4 shows that residual inter-algorithm biases are present when the coefficients are applied to observations, and that these are appropriately addressed by a new method of bias correction. Section 3.5 quantifies the new scheme's performance in simulation against criteria defining a successful approach. Section 3.6 is a discussion, including comments about how the new aspects of our work might apply to other existing and future sensors. For an empirical assessment of the new coefficients' performance in practice, see Embury et al. (2012b). Note that work is ongoing to exploit the satellite overlap periods to homogenize the SST record and to analyse in particular the stability of SST retrieval in detail; these aspects will be reported elsewhere.

Anding and Kauth (1970) proposed that SST could be usefully estimated using a weighted combination of two brightness temperatures (BTs) observed from space. The observations are required to be at two infrared wavelengths that have significant atmospheric transmission and are differentially attenuated by water vapour in passing through the atmosphere. These two conditions mean that the difference between the SST and one of the BTs is approximately linear in the difference between the BTs, both differences being dependent on the atmosphere's variable water vapour content. The approximations underlying this near-linearity were well described by Deschamps and Phulpin (1980). Letting the two BTs be y_1 and y_2, it follows that SST, $\hat{x}$, can be estimated by an expression:

$$\hat{x} = a_0 + a_1 y_1 + a_2 (y_1 - y_2) \tag{3.1}$$

where a_0, a_1 and a_2 are retrieval coefficients.

Wavelengths between about 10 and 13 μm have relatively high atmospheric transmission, and the two BTs are usually observed within this "window", using channels nominally centred on 11 and 12 μm. The above equation is then known as the "split window" retrieval. Barton (1995) provided a review of variants of the split window algorithm, including comprehensive references and a comparison of nineteen published sets of Advanced Very High Resolution Radiometer (AVHRR) coefficients. Coefficients have to be defined for each instrument in a series, since the spectral response functions of nominally similar channels vary between instruments. Operational AVHRR coefficients have traditionally been defined by empirical regression against in situ observations, such as drifting buoys.

A number of elaborations of the basic equation above are now discussed thematically.

Viewing angle. To account for the variation of atmospheric transmission with satellite view zenith angle, θ, across the instrument swath, coefficients need to vary with zenith angle. For AVHRR, the usual approach has been to formulate the i^{th} coefficient as $a_i = a_{i0} + a_{i1} S$ where $S = (\sec(\theta) - 1)$. Often authors have restricted angle dependence to one or a few terms (e.g. Walton et al., 1998), equivalent to forcing a_{i1} to be zero for selected i. Parameterisation with respect to S is convenient for coefficients based on empirical regression to in situ observations, where the satellite zenith angles of the matches cannot be prescribed. The angular variation of S only approximates the additional atmospheric path for off-nadir views. When basing coefficients on radiative transfer, it is useful to interpolate with respect to S between coefficients derived for the simulated view angles (e.g. Merchant et al., 1999).

Additional channels and views. BTs observed around 3.7 to 3.9 µm are useful for SST estimation for night-time scenes (when solar-reflected radiance is absent). Two properties of these channels help: the atmospheric transmission is closer to 1.0 in this window than in the split-window (except for extremely dry atmospheres); and the extreme dependence of radiance on temperature (around T^{14}) reduces the BT impact of atmospheric constituents and "contaminants" such as unscreened sub-pixel clouds. Inclusion of this additional BT has been done using various equations, most of which can be re-expressed as:

$$\hat{x} = a_0 + \sum_i (a_{i0} + a_{i1} S) y_i \tag{3.2}$$

Again, additional contraints are often imposed, such as some a_{i1} being set to zero, certain pairs of coefficients being equal and of opposite sign (e.g. Li et al., 2001). These additional constraints usually express the underlying principle of differential absorption that allows SST to be estimated. But such additional constraints are unnecessary (except to reduce the degrees of freedom being fitted when the number of matches to in situ data is very limited). The ATSRs observe each point on the sea surface in two views, and the generalisation applied by Závody et al. (1995) is simply to let the index i represent each channel-view combination. For example, for an ATSR dual-view three-channel algorithm, i in Eq. 3.2 runs from 1 to 6, indexing the nadir and forward view of each of the channels at 3.7, 11 and 12 µm.

Non-linear terms. The residuals (retrieved minus actual SSTs) of the purely linear estimators represented by equations 3.1 and 3.2 usually display coherent systematic variations geographically, and also with respect to parameters such as latitude, atmospheric total column water vapour (TCWV), and BT differences. This has prompted trials of various non-linear estimators. This may be achieved by having different coefficients for bands of different latitude (e.g. Závody et al., 1995), BT difference, (e.g. Walton et al., 1998) or atmospheric water vapour (e.g. Barton, 1998). Another approach is to include a coefficient multiplying a prior SST estimate Walton et al. (1998) or prior water vapour Emery et al. (1994). The non-linear SST form by Walton et al. (1998) is used in the AVHRR Pathfinder SST project (Kilpatrick et al., 2001). Quadratic terms and other non-linear functional forms have been tried (e.g. Walton, 1988; Emery et al., 1994). These offer limited benefit to retrieval accuracy because simple

functional forms do not reflect the underlying origin of the systematic errors. Merchant et al. (2006b) explained the origins of systematic errors in terms of the non-linearity in atmospheric radiative transfer and the geographical variations in the geophysical factors that affect BTs.

Aerosols. The major volcanic eruptions of El Chichon (Mexico, April 1982), Pinatubo (Phillippines, June 1991) and Cerro Hudson (Chile, August 1991) injected sulphurous material into the stratosphere, leading to the presence of a sulphuric acid aerosol layer that persisted in the stratosphere for upward of one year (e.g. Baran and Foot, 1994). The additional atmospheric attenuation from stratospheric aerosol tends to introduce bias into SST estimators not designed for such conditions. Walton (1985) noted that the sensitivity to stratospheric aerosol varies with the coefficients used, and proposed a less sensitive form for AVHRR SSTs for the post-El-Chichon period. Merchant et al. (1999) pointed out in the context of ATSR that the impact on BTs of stratospheric aerosol can be simulated and coefficients chosen such that the SST estimator has strictly zero sensitivity to those simulated BT impacts. This proved effective for ATSR-1 dual-view retrievals, and was implemented operationally. The technique is inapplicable to split-window retrievals, since at least three BTs are required to correct for both atmospheric gas absorption and stratospheric aerosol effects independently. Tropospheric aerosol variability can also affect estimated SSTs, and an approach to correction for AVHRR was proposed by Nalli and Stowe (2002). For the Spinning Enhanced Visible and Infra-Red Imager (SEVIRI), Merchant et al. (2006a) established a correction for Saharan Dust, where presence can be identified with the help of the channel at 8.7 μm on that sensor. Noyes et al. [this issue] present a Saharan Dust Index for the ATSR instruments.

Radiative transfer modelling. Empirical regression of BTs to in situ measurements has been widely used to define coefficients. Others base coefficients on simulations of BTs by radiative transfer modelling for a range of SST and atmospheric conditions. The attempt by Závody et al. (1995) was hindered by an inadequate spectroscopic data describing the water vapour continuum absorption of the atmosphere (Merchant et al., 1999). With current spectroscopic knowledge, however, simulation-based coefficients are routinely made operational with only modest empirical adjustment to the overall offset coefficient, a_0 (Merchant and Le Borgne, 2004), for generation of products from the Ocean and Sea Ice Satellite Application Facility (France) and some geostationary SST products from the National Oceanic and Atmospheric Administration (US).

Limitations of coefficient-based retrievals of SST have been identified related to sensitivity of the SSTs to unusual atmospheric conditions, such as dry or humid layers (Minnett, 1986). By combining empirical and simulation techniques, Merchant et al. (2009) characterised the non-linear SST performance and quantified its intrinsic regional bias and sensitivity to water vapour and SST.

Significant potential for improved SST using optimal estimation has been demonstrated in a study on AVHRR (Merchant et al., 2008b). However, with the exception of the nadir-only split-window (N2) retrieval, with ATSR we have found so far that SSTs based on optimal estimation have larger systematic biases (~0.2 K) than SSTs based on the new coefficient design reported here. Thus, the coefficient based approach to retrieval continues to be pursued within the ATSR Reprocessing for Climate (ARC;

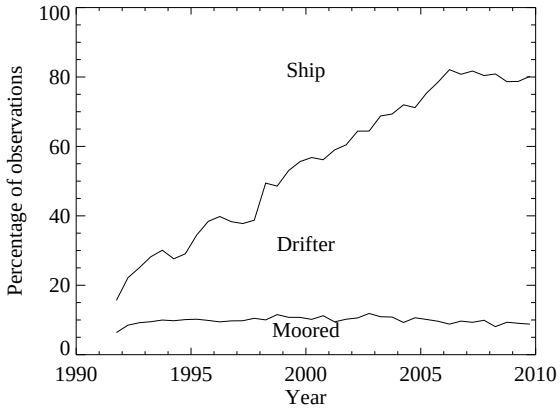

Figure 3.1: Change in observation types present in ATSR-in situ match-up databases (Embury et al., 2012b). The proportion of moored buoy observations is that below the lower line. The proportion of drifting buoy observations is that between the lower and upper line, the remainder above the upper line is the proportion of ship observations. These proportions agree closely to those found in a full in situ record. (J. Kennedy, personal communication).

Merchant et al., 2008c) project.

ATSR retrieval coefficients have from the start been defined by radiative transfer modelling rather than empirically (Závody et al., 1995). This means that the SSTs obtained are independent of in situ observations. This is important for climate research because ocean mixed layer temperature plays a crucial role in decadal climate (e.g. Hegerl and Bindoff, 2005), and SST is the key parameter to quantify the marine component of global warming (e.g. Rayner et al., 2003). It is beneficial to have an independent record of SST to increase confidence in estimates of change in marine climate in recent decades, particularly since the balance of types of in situ SST observation have evolved rapidly since 1990 (see Fig. 3.1). Our aim for the new ATSR coefficients is therefore to preserve independence from the in situ record by basing the coefficients on radiative transfer, while obtaining SSTs that are usable for climate analyses. The aspiration is that global and regional biases should be less than 0.1 K in magnitude, with stability of observation of ~0.05 K/decade (Merchant et al., 2008c).

3.2 ATSR SST Retrieval

3.2.1 Background

SSTs are routinely derived from ATSR observations using four combinations of cloud-screened BTs. The nadir view (encompassing zenith angles from about 0 deg to 22 deg with pixel ground resolution of about 1 km) is used alone to deliver nadir-view two-channel SSTs (i.e., split-window retrieval using 11 and 12 μm BTs). Hereafter, these are designated “N2” SSTs. N2 SSTs are valid for night and day scenes. Nadir-view three-channel SSTs (N3 SSTs) additionally use the 3.7 μm channel, and are therefore valid only when solar irradiance in the shorter wavelength channel is negligible. Forward-view observations

are obtained at satellite zenith angles around 55 deg, and have a pixel ground resolution of about 1.8 by 3 km. However, both forward and nadir pixels are regridded on to a common ~1 km resolution grid by a nearest neighbour method, and are in practice treated as observations of the same location on the ground at different view angles. The views are used together for "dual-view" retrievals. Analogously to the nadir case, both dual-view two-channel (D2) and dual-view three-channel (D3) SST coefficient sets are defined.

Having two views of different atmospheric path provides additional information about the effect of the atmosphere, allowing greater robustness to atmospheric aerosols (Merchant and Harris, 1999) and smaller coefficient-related systematic errors (Merchant et al., 2006b). However, the forward view channels also add radiometric noise, and sampling error arising from the different pixel projections on the sea surface. In simulations where these error contributions are neglected, dual-view SSTs are always more accurate than nadir-view SSTs. In practice this is true for the N2/D2 retrievals used during the day; but the difference is less marked for N3/D3 available at night, except that dual-view retrievals are more reliable when atmospheric aerosols are significant (Noyes et al., 2006).

In this paper, we also introduce a further channel combination in addition to the long-standing N2, N3, D2 and D3 combinations. We call the new combination "D2*", and it consists of the 3.7 and 11 µm channels in dual view. The D2* algorithm allows us to deal with uncertainty in the calibration of the 12 µm of AATSR, as discussed in subsection 3.4.2. In simulation, the behaviour of D2* is almost indistinguishable from that of D3, with simulated residuals differing by <0.01 K in general. Therefore, the simulation results presented in this section for D3 reflect the performance of D2* very closely.

3.2.2 Previous ATSR coefficients

It is useful to use vector-matrix notation hereafter. Coefficients are written as a_0 (offset coefficient) and $\mathbf{a}$ (weighting coefficients listed in a column vector). BTs are listed in the observation vector $\mathbf{y}$, and the linear estimator for SST (Eq. 3.2) is then written as:

$$\hat{x} = a_0 + \mathbf{a}^{\mathrm{T}}\mathbf{y} \tag{3.3}$$

where the summation explicit in Eq. 3.2 is achieved by the matrix multiplication of the coefficient and observations vectors. The observation vector comprises six BTs (three channels and two views), and all but the D3 SST estimator have some elements of $\mathbf{a}$ equal to zero (corresponding to the unused channels and/or view).

Due to the conical scan method used by the ATSR instruments [see for example, Llewellyn-Jones et al, this issue] the difference in zenith angle between forward and nadir views varies across the 512-pixel instrument swath. If the satellite is flying nominally, the zenith angles are symmetric about the ground track, and the greatest difference is found at the swath centre and lowest contrast at the edge – see Table 3.1. Závody et al. (1995) captured this effect by banding the swath by number-of-pixels-from-the-swath-centre in five bands, each with their own coefficients. However, discontinuities

Position	Nadir angle	Forward angle	Pixel index
Centre	0.04	55.59	±1
Edge	21.73	53.01	±256

Table 3.1: Nominal ATSR viewing geometery. Pixel index indicates across-track offset from sub-satellite position.

at band boundaries were not always negligible using this method. Merchant et al. (1999) introduced the approach used routinely for the ATSRs more recently. They specified retrieval coefficients at the centre and edge of the swath (using a fixed nominal flight configuration). Between the centre and edge, coefficients for 38 across-track bands were found by interpolating between the "centre" and "edge" cases linearly with respect to the nominal atmospheric path for the nadir view.

Operational dual-view coefficients have also been specified to be robust to two modes of stratospheric volcanic aerosol, "aged" and "background", following Merchant et al. (1999).

3.2.3 Derivation of coefficients

The original ATSR coefficients were generated using standard least-squares regression of simulated BTs to the SSTs used in their simulation. The simulations were for atmospheric conditions defined by a set of radiosonde profiles. The regression equation was (Závody et al., 1995):

$$\begin{aligned} \mathbf{a} &= (\mathbf{S_{yy}} + \mathbf{S}_\varepsilon)^{-1}\mathbf{S}_{x\mathbf{y}} = \mathbf{S'^{-1}_{yy}}\mathbf{S}_{x\mathbf{y}} \\ a_0 &= \bar{x} - \mathbf{a}^{\mathrm{T}}\bar{\mathbf{y}} \end{aligned} \tag{3.4}$$

using the nomenclature of Merchant et al. (1999), where x and $\mathbf{y}$ are the SSTs and simulated BTs respectively from the training dataset, with an over bar representing the mean value. $\mathbf{S}_\varepsilon$ is the covariance matrix representing radiometric noise, taken as diagonal implying independence between channels. $\mathbf{S_{yy}}$ is the covariance matrix of (simulated) BTs and $\mathbf{S}_{x\mathbf{y}}$ the covariance vector of SST and BT, given by:

$$\begin{aligned} \mathbf{S_{yy}} &= \overline{\mathbf{y}\mathbf{y}^{\mathrm{T}}} - \bar{\mathbf{y}}\bar{\mathbf{y}}^{\mathrm{T}} \\ \mathbf{S}_{x\mathbf{y}} &= \overline{x\mathbf{y}^{\mathrm{T}}} - \bar{x}\bar{\mathbf{y}}^{\mathrm{T}} \end{aligned} \tag{3.5}$$

finally, $\mathbf{S'_{yy}} = \mathbf{S_{yy}} + \mathbf{S}_\varepsilon$

This method was adapted by Merchant et al. (1999) to produce stratospheric-aerosol robust coefficients. Here, the effects of stratospheric aerosol were characterised as altering the top-of-atmosphere BTs from $\mathbf{y}$ to $\mathbf{y} - c\mathbf{k}$, where the vector $\mathbf{k}$ represents the relative magnitude of deficit in BT for different channels/views, and c is proportional to the aerosol optical depth; $\mathbf{k}$ is evaluated by radiative transfer modelling at the centre and edge of the swath. The retrieved SST is robust (i.e. insensitive) to the simulated stratospheric aerosol impacts if $\mathbf{a}^{\mathrm{T}}\mathbf{k} = 0$, since the retrieved SST is then unchanged whether the observations consist of $\mathbf{y}$ or $\mathbf{y} - c\mathbf{k}$. This property of coefficients can be

achieved by modifying Eq. 3.4 to:

$$\mathbf{a} = \mathbf{S}'^{-1}_{\mathbf{yy}} \left[\mathbf{S}_{\mathbf{xy}} - \mathbf{K} \left(\mathbf{K}^{\mathrm{T}} \mathbf{S}'^{-1}_{\mathbf{yy}} \mathbf{K} \right)^{-1} \mathbf{K}^{\mathrm{T}} \mathbf{S}'^{-1}_{\mathbf{yy}} \mathbf{S}_{\mathbf{xy}} \right] \tag{3.6}$$

where **K** is a matrix of all the vectors **k** to which the retrieval is chosen to be robust. Assuming any **k** (or any linear combination of **k**s) within **K** accurately describes their BT impact, the retrieved SSTs are negligibly affected by the presence of stratospheric aerosols. Merchant and Harris (1999) established that this approach works well in practice for D2 and D3 SSTs for ATSR-1. These estimators have enough degrees of freedom (four and six observations respectively) to provide an atmospheric correction for the non-aerosol atmospheric effects and one or two modes of stratospheric aerosol, without significantly increased retrieval noise. In principle, N3 SSTs can also be robust to one mode of aerosol, i.e., a single **k**, at the expense of increased random retrieval error (greater amplification of radiometric noise); but this depends more critically on the accuracy of **k** than the dual view case, is not as effective in practice (Merchant and Le Borgne, 2001), and was not routinely implemented. There is no possibility of aerosol robustness by this method for split-window (N2) SSTs. Aerosol-robust dual-view coefficients have been defined for all three sensors. This proved unnecessary for ATSR-2 in that there were no stratospheric aerosol episodes during its mission. AATSR is prepared to continue operational delivery of good SSTs should a major volcanic eruption cause elevated stratospheric aerosol levels during its lifetime.

3.3 New ATSR retrieval scheme

3.3.1 Atmospheric composition

We now describe the new retrieval coefficient scheme developed for re-processing of the ATSRs within ARC (Merchant et al., 2008c). The coefficients are based on comprehensively updated radiative transfer simulations, as fully described in Embury et al. (2012a). Briefly, we simulate ATSR BTs for a comprehensive set of surface and atmospheric conditions (Chevallier, 2002), derived from ERA-40 (Uppala et al., 2005). The range of meteorological variability in this data set is assumed sufficient to support the determination of coefficients for any year and season. Simulations are performed using a line-by-line radiative transfer model at high resolution exploiting updated spectroscopic data and improved surface emissivity. In addition, we have refined the way in which the coefficients are stratified with respect to various factors.

The first factor relates to the secular change in atmospheric trace gases (not including water vapour) over the course of the ATSR missions. Our assumptions about how atmospheric composition varies are detailed in Embury et al. (2012a), along with the relative impacts of trace gases. We undertook radiative transfer modelling for the trace gas conditions prevailing at intervals of 4 years (1991, 1995, through to 2007). Profiles are distributed throughout the year to capture seasonal variations in trace

gases (i.e. nitric acid), in order that coefficients respond to an appropriate average distribution and reflect seasonal variability to the extent possible. The average trace gas impact has a trend in time, and to avoid a corresponding trend artefact in retrievals, the coefficients are interpolated in time across each four-year interval. July 1st is taken as the mid-point of each coefficient set for the purposes of the interpolation. In the ARC data processing chain, this time interpolation is done once per orbit file, using the start time of the orbit, which is more than adequately precise.

Not accounting for trace gas changes would lead to a relative trend between simulations and observations of ~0.01 K/decade in 3.7 and 11 µm BTs. This drift would be slightly higher during the 1990s and slightly lower in 2000s. When these BTs impacts are applied to ATSR SST retrievals coefficients, the effects are magnified by the coefficients for nadir-only retrievals, but reduced by the dual-view coefficients. The worst affected would be the N2 retrieval which would show trends ~0.05 K/decade, while the N3 drift is only half the size. Dual-view D2 and D3 are comparatively robust to the changes in trace gases: trends would be ~0.005 K/decade. Given our target stability, 0.05 K/decade, it is important to account for these trace gas trends, at least for nadir-only SSTs.

3.3.2 Viewing geometry

The variation of coefficients across the instrument swath has been refined. The previous approach (see subsection 3.2.2) assumes (i) that forward-view path length varies in proportion to the nadir path length across the instrument swath, and (ii) that actual satellite zenith angles across the track agree closely with the nominal values. These approximations cause non-negligible errors.

Figure 3.2 shows the bias in D3 SST arising from interpolating coefficients across the swath with respect to nadir-path length. This has been calculated by interpolating the radiative transfer output to the correct nadir and forward zenith angles, according to the nominal viewing geometry, and applying the centre/edge coefficients to the simulated BTs. The same dataset is used for testing the coefficients as is used for defining them (i.e. the training dataset), so the retrieval bias is expected to be zero at centre and edge of swath. However, for the majority of the swath there is a positive bias reaching 0.05 K around 64 pixels from the swath edge. D2 biases, not shown, follow the same pattern but with ~30 % greater magnitude.

The reason for this 'm' shaped bias pattern is the different zenith angle sensitivity of the two coefficient sets. (There is also a small contribution from the assumption that forward view path length varies linearly with the nadir; however, this contributes <0.01 K to the bias.) If the centre coefficients were applied across the whole swath the bias would be zero at swath centre increasing to 0.9 K at the swath edge; however, applying the edge coefficients to the whole swath gives −0.65 K bias at swath centre decreasing to zero at the edge. Interpolating between the two therefore gives non-zero bias between the two extremes. This effect is only noticeable in the dual-view retrievals as they are much more sensitive to the viewing geometry. For N3 retrieval there is a bias of −0.085 K for centre-coefficients applied to edge BTs, and +0.088 K for edge-coefficients applied to centre-BTs.

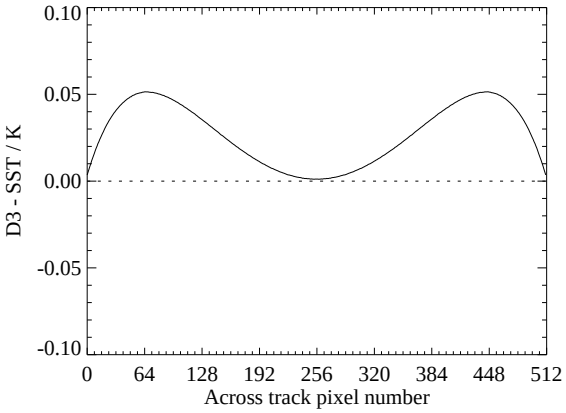

Figure 3.2: D3 retrieval bias introduced through interpolating between centre and edge coefficients for the nominal viewing geometry case, as previously done for SST estimation.

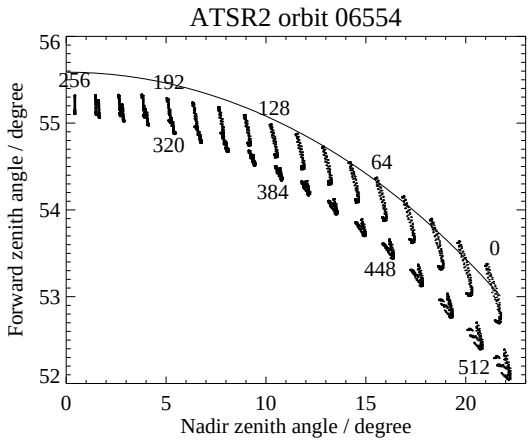

Figure 3.3: Forward zenith angle against nadir for ATSR-2 orbit 06554 (00:01 22/07/1996). Each cluster of points represents the view angles for one across-track position sampled along the orbit. The across-track pixel position is labelled for every fourth cluster. Solid line shows nominal ATSR-2 viewing geometry.

Thus, interpolation errors affecting the nadir retrievals are two orders of magnitude smaller than for the dual-view retrievals.

Variation in instrument pointing means that the viewing geometry can in practice be asymmetric across the instrument swath, and orbital ellipticity and orbit manoeuvres modify the edge-of-swath zenith angles. Figure 3.3 shows the difference between the nominal viewing geometry (solid line) and the actual geometry at 32 equally spaced positions across swath for one randomly selected ATSR-2 orbit.

The effects of these viewing angle variations are simulated as before; simulated BTs from the training dataset are generated for each of the observed nadir/forward viewing angles, centre/edge coefficients are then applied to the BTs to determine the retrieval bias. Figure 3.4 shows the resulting D3 bias as a function of across track pixel number and latitude (D2 biases are ~30 % larger). The 'm'

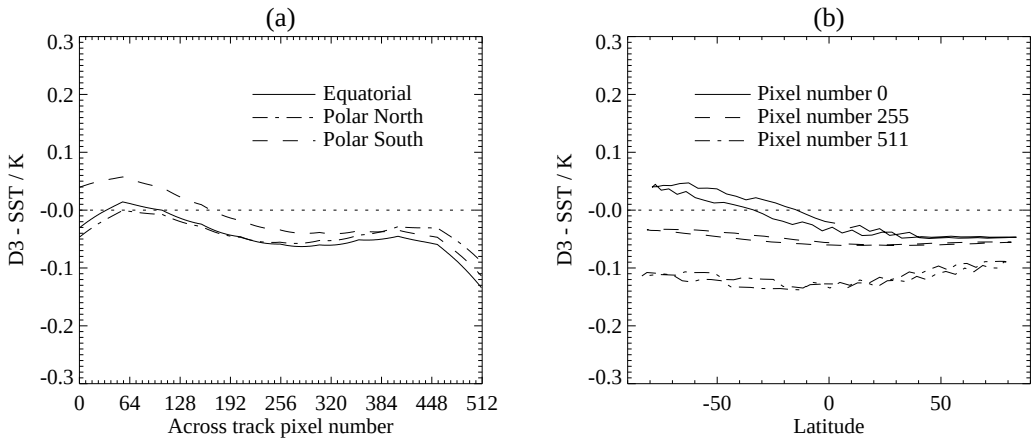

Figure 3.4: D3 retrieval bias introduced through interpolating between centre and edge coefficients for actual viewing geometry. Biases calculated by interpolating BTs to viewing geometry from Fig. 3.3 and applying centre/edge retrieval. Dataset used is constant across all pixel locations/latitudes – i.e. only viewing geometry is variable.

shaped bias seen in Fig. 3.2 is still present, but modified by an extra across track and orbitally varying component. These additional biases are primarily due to the forward view angle not matching the assumption implicit in the coefficients. The D3 retrieval coefficients have a sensitivity of ~0.1 K/degree to the forward view angle, as the actual forward view angle varies by up to a degree from nominal across the swath we find slowly varying systematic biases ~0.1 K in the retrieval.

The new approach taken for ARC is to generate coefficients for a range of different nadir/forward zenith angle combinations and then bi-linearly interpolate between these based on the satellite zenith angles reported in the level 1 ATSR product. The viewing angles chosen correspond to 5 nadir path lengths (1.00, 1.02, 1.04, 1.06, and 1.08) and 6 forward path lengths (1.60, 1.64, 1.68, 1.72, 1.76, 1.80) giving a total of 30 different coefficient sets in the lookup table, for a given epoch and type of retrieval.

The new coefficient interpolation method is applied to the simulated BTs as before and resulting bias shown in Fig. 3.5 – note that the scale has been increased a factor 10 compared with Fig. 3.4. Across swath and intra-orbit variations are now measured in thousandths of a kelvin, about two orders of magnitude smaller than previously. Interpolating between coefficients still produces warm biases analogous to the 'm' shape seen with centre/edge coefficients, but with greatly reduced amplitude due to the larger number of interpolation points. An average warm bias ~0.002 K is seen for this orbit; however this is well within target accuracy.

3.3.3 Water vapour banding

Coefficient-based retrievals can be subject to systematic errors, as discussed in Merchant et al. (2006b). These errors are related to the non-linearity of radiative transfer and the geographical variations in the factors determining BTs. The main factors determining BT in low-aerosol conditions are: SST, TCWV, broad vertical distribution of water vapour, air-sea temperature difference and tropospheric

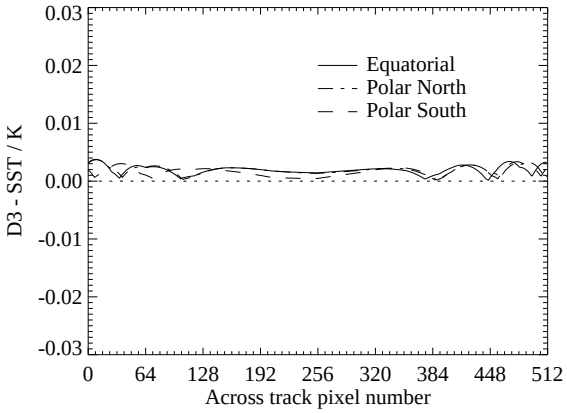

Figure 3.5: D3 retrieval bias from new interpolation scheme. Note scale is increased ×10 compared to Fig 3.4.

Retrieval	N2	N3	D2	D3
Global	0.39	0.11	0.21	0.07
Banded	0.28	0.07	0.11	0.06

Table 3.2: Standard deviation of simulated residuals (see text) for centre-swath AATSR coefficients.

lapse rate. Each of these has a distinct geographical distribution, and varies locally in time. Note also that there are usually more factors than independent observations (channels/views). Therefore, only the dominant aspects of this variability and non-linearity can be fitted by the regression procedure, and lesser aspects that are not fitted give rise to systematic errors. These include geographical biases and non-linear dependences on TCWV, etc. These errors are intrinsic to the nature of the retrieval, not a deficiency in the radiative transfer simulations, and therefore can be modelled and explored using the simulations on which the coefficients are based. To do this, the derived coefficients are applied to the simulated BTs and the difference between the resulting "simulated retrieved SST" from the true SST (the SST used in the simulation) is the retrieval error, or "simulated residual". Because this is all in simulation, the simulated residual does not included the effects of errors in the radiative transfer model or instrumental noise; it purely describes intrinsic limitations of the retrieval scheme.

Table 3.2 shows in the top line of data the standard deviation (SD) of the simulated residuals for AATSR for each type of retrieval. (The global mean residuals are zero by the properties of unweighted least squares regression.) The single view split-window N2 retrieval has the largest SD (0.39 K) while its dual-view equivalent is 0.21 K. This reflects the extra information brought to the retrieval by having two views. The N3 and D3 retrievals have substantially lower retrieval error due to the use of the 3.7 µm channel.

While the global mean residual is zero in each case, when looking at subsets of the retrievals, non-zero biases can be found – these are the systematic errors intrinsic to the retrieval scheme. The systematic errors have complex geographical distributions as shown in Fig. 3.6. As expected from the retrieval errors from Table 3.2, the N2 retrieval has the largest geographical variations in bias (ranging

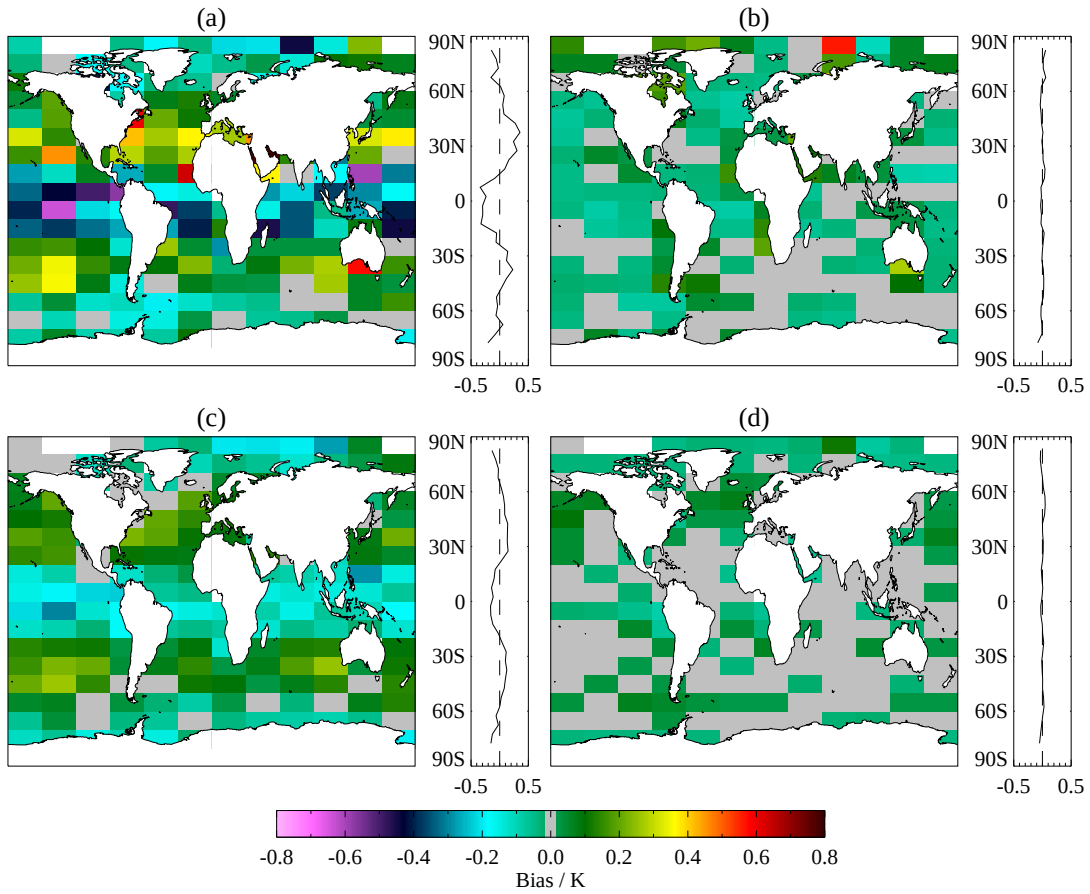

Figure 3.6: Geographical variation in intrinsic retrieval bias due to use of global retrieval coefficients. Coefficients for different channel combinations are derived from radiative transfer simulations for a range of atmospheric profiles (see main text). These coefficients are then applied to the same simulations to obtain simulated retrieved SSTs, from which the input-profile SSTs are subtracted. In this way, the systematic regional differences that exist even for "perfect" coefficients can be simulated (Merchant et al., 2006b). (a) N2, (b) N3, (c) D2, and (d) D3.

from −0.64 to 0.78). The D2 retrievals biases range from −0.3 to 0.26 and resemble those previously seen for operational AATSR D2 retrievals (Merchant et al., 2006b). The N3/D3 retrievals have only minimal systematic variation due to the properties of the 3.7 µm channel discussed above (range is from −0.1 to 0.1, with the exception of a single cell where N3 is 0.5 K warm).

The systematic variation with latitude has been previously noted and explained (Merchant et al., 2006b), and a "latitudinal offset correction" has been implemented for the operational D2 retrieval (Birks, 2006). Meanwhile, operational N2 and N3 are banded in three broad latitude zones, following Závody et al. (1995). However, neither of these approaches can improve the longitudinal structure that is also evident in Fig. 3.6.

Prior estimates of TCWV from numerical weather prediction are available within the ARC processing chain. (These are obtained from the European Centre for Medium-range Weather Forecasting, ECMWF. Up to 2002, the "ERA-40" ECMWF Re-Analysis is used (Uppala et al., 2005), and thereafter the TCWV is obtained from operational analyses.) The dependence on TCWV is shown, for the D2 retrieval,

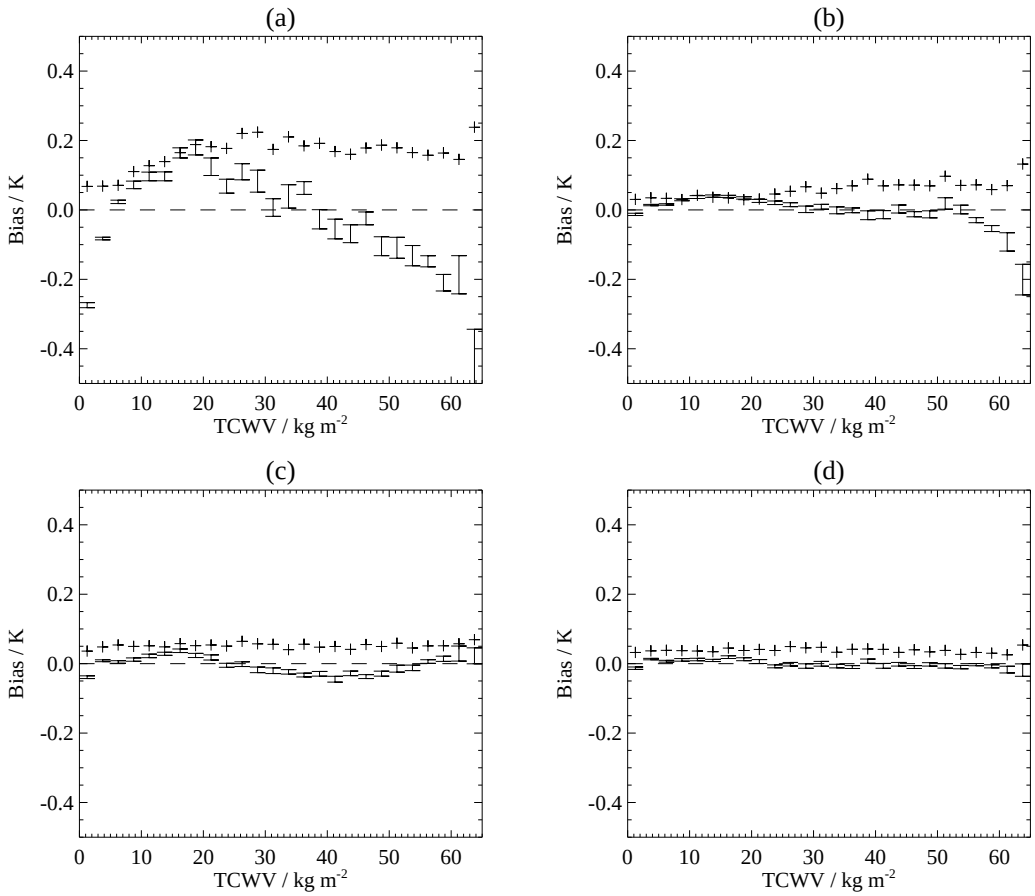

Figure 3.7: Intrinsic retrieval bias as a function of TCWV for (a) D2 global coefficients, (b) D2 water vapour banded coefficients, (c) D3 global coefficients, and (d) D3 water vapour banded coefficients. By intrinsic, we mean that the bias arises as a result of the retrieval design, even for perfectly specified coefficients. Error bars show mean and standard error for each bin. Crosses indicate standard deviation within each bin.

in Fig. 3.7a. Systematic biases, cold for both very wet and very dry atmospheres, are consistent with the latitudinal bias structure seen in Fig. 3.6c where the retrieval is cold biased for both equatorial and polar regions. As water vapour, unlike latitude, has a direct effect on the simulated BTs it is a more appropriate choice for a variable to use to inform the SST retrieval.

As the dependence on TCWV shown above is not accurately captured by a quadratic fit, we follow the proposal of Barton (1998) and band coefficients by TCWV. We stratify the coefficients using overlapping bands of TCWV (in the simulation profile set) that are 20 kg m^{-2} wide and stepped by 5 kg m^{-2} intervals – i.e. the bands are [0,10], [0,15], [0,20], [5,25], ... [50,70]. SD of simulated residuals using TCWV banded coefficients is shown in Table 3.2. Compared with the global coefficients, there has been a significant improvement in the retrieval error for the N2/D2 retrievals; standard deviations of the simulated residuals are now 0.28 and 0.11 as opposed to 0.39 and 0.21 using the global coefficients. N3/D3 retrievals, which already had low SDs, are only slightly improved with banded coefficients. Figure 3.7b shows the D2 retrieval bias against TCWV; the cold bias for extremely dry atmospheres has been eliminated and the warm bias for typical TCWV values (15 to 35 kg m^{-2}) has

been reduced from a maximum of 0.2 K to <0.05 K. For the wettest atmospheres (TCWV~60 kg m^{-2}) a cold bias remains, but is reduced in both magnitude and extent compared to the global coefficient case. Although the bias could be further improved through the use of narrower water vapour bands, this would increase the dependence on the prior TCWV (see section 3.5), which is undesirable.

Figure 3.7c and d illustrate the relative absence of TCWV-dependent structure in simulated D3 retrievals, with or without banding. This does not preclude the possibility of TCWV dependence in D3 retrievals when applied to observations; it does demonstrate that any such behaviour is not an intrinsic limitation of the coefficient design.

Figure 3.8 shows the geographical distribution of retrieval bias for the banded coefficients. The N3/D3 retrieval biases remain comparable to the global coefficients shown in Fig. 3.6 with most areas showing bias <0.02 K in magnitude. The N2/D2 have been greatly improved. The regional structure in the D2 bias is improved: the number of cells with bias exceeding 0.1 K is reduced from 10% to 4%, these remaining cells all being coastal with fewer profiles, and therefore with increased standard error in the estimate of the mean. The N2 retrieval has 30% of cells with magnitude of retrieval bias >0.1 K (compared to 60% for the global coefficients). The geographical structure has been moderately reduced, the remaining geographical features being warm biases in some coastal regions and a cold equatorial bias.

3.3.4 Radiometric noise

The simulated BTs used to generate the retrieval coefficients, unlike actual observations, do not include any effects of instrumental noise. This is accounted for in Eq. 3.6 through the $\mathbf{S}_\varepsilon$ term – the covariance matrix of the radiometric noise, or noise equivalent differential temperature (NEdT) – which is equivalent to adding a Gaussian noise component of the correct magnitude to the simulated BTs. Accounting for noise in this way has the effect of limiting the magnitude of the components of the coefficient vector, $\mathbf{a}$, since the variance of the radiometric error in the SST retrieval is:

$$\sigma_\varepsilon^2 = \mathbf{a}^{\mathrm{T}} \mathbf{S}_\varepsilon \mathbf{a} \tag{3.7}$$

However, minimising σ_ε comes at a price – the variance of the "simulated residuals," $\sigma_x^2 = \mathrm{Var}(\hat{x} - x)$, is increased. In fact Eq. 3.6 minimises the total error given by the sum of these two terms. This ensures that each SST retrieval has the lowest possible error variance.

However, we do not necessarily want the "lowest possible" error because not all types of errors are of equivalent importance to climate applications. The radiometric error is random and uncorrelated from one pixel to the next. The intrinsic retrieval errors (as determined by simulating residuals in the absence of noise) are systematic; we have already reduced part of these systematic errors through the use of water vapour banded coefficients (see above), but systematic responses to the atmospheric state are still present. These can be considered pseudo-random: they are correlated on synoptic to seasonal space-time scales, but are often misinterpreted as random scatter in validation exercises.

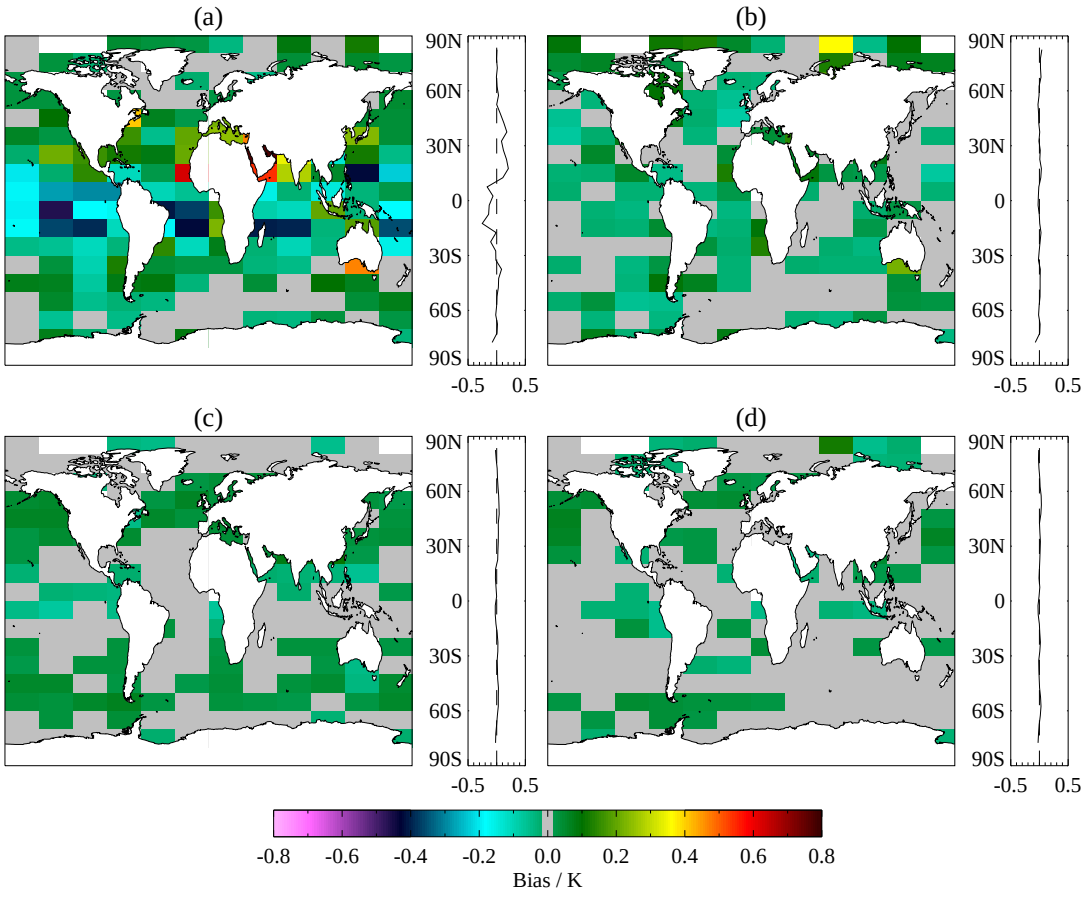

Figure 3.8: As Fig 3.6 using TCWV banded coefficients.

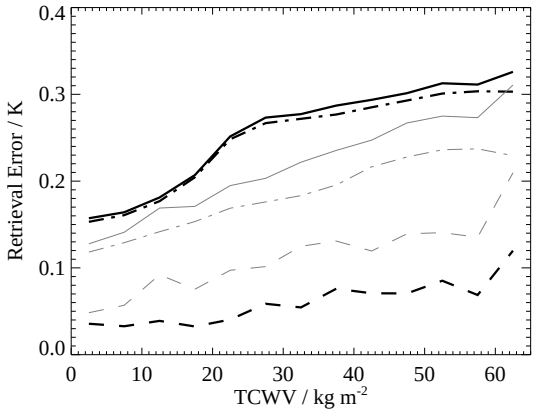

Figure 3.9: Radiometric (dot-dashed) and pseudo-random (dashed) components of simulated total retrieval error (solid) for D2 coefficients designed using actual instrument noise levels (grey, thin) and a nominal noise of 0.01 K (black, thick)

For some applications, e.g. measuring ocean temperature gradients, it is more important to minimise the radiometric error than the the pseudo-random error. However, in the context of climate and spatially averaged SST products, radiometric error is less relevant because over a sufficient number of observations it will average out to zero, while the pseudo-random error, being correlated between pixels, will remain. To generate the retrieval coefficients we use noise of 0.01 K in each channel. This compares to actual radiometric noise in pixel BTs that is typically of order 0.05 K. Figure 3.9 shows the consequence of this compared to coefficients generated using the actual instrument NEdT in terms of the random (σ_ε), pseudo-random (σ_x), and "total" errors for the D2 retrieval. The total retrieval error (solid line) is always greater for the minimal-noise coefficients (black) than the actual-noise coefficients (grey), but this is mostly due to the increase in sensitivity to the radiometric noise (dashed+dotted line). The pseudo-random error (dashed line) is significantly reduced for the minimum-noise coefficients. Even for a spatially averaged product at 0.05° or ~5 km resolution, the radiometric noise will be reduced by up to a factor of 5 meaning that the pseudo-random error is dominant.

In principle we should set $\mathbf{S}_\varepsilon = \mathbf{0}$, i.e. assume zero instrumental noise, when generating the retrieval coefficients since our priority is to minimize pseudo-random errors, rather than random error, for climate applications. In practice we find there is no benefit – the decrease in pseudo-random retrieval error is negligible ~0.001 K. Moreover, there is also a disadvantage in doing so. With a zero-noise assumption we found that, in some TCWV bands, the magnitude of some coefficients became unduly large. This may reflect numerical problems in determining the inverse in Eq. 3.6. Large coefficients have an adverse effect on biases if the radiative transfer simulations are imperfect – which inevitably they are. Large coefficients tend to amplify systematic error from errors in the forward modelling process. So, it is better to condition the matrix inversion (Eq. 3.6), and a nominal amount of noise of 0.01 K is sufficient to do so.

3.3.5 Practical implementation

To use the coefficients defined as described in this section, the following steps are required during processing. The first step is to select the coefficients of the correct sensor and interpolate the coefficients in time between the years (at four year intervals) of radiative transfer simulations. This need be done only once per orbit or image extract, using the start time of acquisition. Step two is to obtain, for the image pixel in question, the TCWV and nadir and forward-view zenith angles. The TCWV interpolation may involve space and time interpolation from auxiliary numerical weather prediction fields. The viewing angles can be interpolated from information available in the satellite data file. The third step is interpolation of coefficients with respect to viewing angles and TCWV, an appropriate approach being bilinear interpolation with respect to nadir and forward-view zenith angle, followed by linear interpolation with respect to the TCWV at the centre of the two nearest TCWV bands. This leads to the interpolated coefficients that can, finally, be applied to the brightness temperatures of the pixel.

3.4 Forward modelling error correction

3.4.1 Inter-algorithm comparison

In the previous section, we have not considered the errors inherent in the forward modelling process. The possible errors comprise: (i) inaccuracy in the radiative transfer simulation (both because of the model and because of imperfect characterisation of the instrumental response or calibration); and (ii) systematic error in the distribution of the profiles used as input (unrealistic distribution of water vapour, too much or too little marine aerosol, etc). The most obvious consequence of forward modelling error is that the SST retrievals have a global mean bias (Merchant and Le Borgne, 2004).

The global biases arising from forward modelling error differ between different types of retrieval. This is because the forward modelling errors in general differ between channels, and different types of retrieval give channels different weights. If the forward modelling were perfect, however, the global bias for all types of retrieval would be zero (assuming consistent sampling between the observations and the profiles used to define the coefficients). Therefore, without reference to any in situ SST data, we can infer much about the quality of forward modelling by considering the differences in estimated SST when coefficients are applied to a sample of clear-sky observations.

We use a global sample of night-time clear-sky BTs from the data base of matches of AATSR to drifting buoys used in Embury et al. (2012b). (We emphasize that the in situ SSTs themselves are not used in this analysis, despite being present in the data set.) For each match, the N2, N3 and D2 SSTs are calculated and from these the D3 SST is subtracted to form an inter-algorithm difference for the new coefficients ("ARC coefficients" hereafter). This is also done using the operational AATSR coefficients ("OP coefficients" hereafter).

Inter-algorithm differences for AATSR are shown in Table 3.3 and Fig. 3.10 for both the OP and

Retrieval	N2-D3	N3-D3	D2-D3
OP	$+0.38 \pm 0.42$	-0.01 ± 0.17	-0.20 ± 0.21
ARC	-0.09 ± 0.30	-0.02 ± 0.13	-0.12 ± 0.14

Table 3.3: Inter-algorithm differences (bias ± standard deviation) for operational (OP) and new (ARC) AATSR retrieval coefficients. The difference between OP and ARC (D3) is 0.06 ± 0.07.

ARC coefficients. The operational coefficients are based on older radiative transfer modelling [A Birks, personal communication]. The OP N3 and D3 retrievals are in good agreement, the systematic variation in their difference being between −0.07 and +0.08 K for TCWV between 5 and 50 kg m^{-2} (Fig. 3.10b). The OP N2 and D2 retrievals are both significantly different to the D3 retrievals, and differ from each other by between 0.2 and 0.7 K, depending on TCWV. Fig. 3.10a and Fig. 3.10c show a variation of several tenths of kelvin with TCWV. The OP D2-D3 global difference of about −0.2 K is similar to the D2-D3 differences seen in previous sensors in the mission (e.g. Merchant and Harris, 1999, with reference to ATSR-1). Overall, the OP N3, D2 and D3 agree to within 0.2 K for the most common range of TCWV (15 to 35 kg m^{-2}). Reflecting this good performance, AATSR OP SSTs have been used as SST estimates of reference in some contexts (e.g. Stark et al., 2007).

The new ARC coefficients have improved inter-algorithm consistency. The largest global inter-algorithm difference, ARC D2-D3, is −0.12 K. The ARC N2 and D2 retrievals are ~0.1 K colder than the N3 and D3 retrievals, a relative bias that is therefore associated with the use (or absence) of the 3.7 µm channel. Likewise, the TCWV dependence is up to ~0.1 K for D2 relative to D3, while N3 and D3 show minimal relative difference with respect to TCWV. Figure 3.10d shows the difference between the OP D3 and ARC D3 SSTs – there is no significant water vapour dependence and only a small 0.06 K bias between the two. Improved consistency is also shown by generally reduced standard deviation between algorithms within a given band of TCWV in Fig. 3.10a to 3.10c. Reductions are significant at low TCWV values, and generally modest for the most humid profiles, reflecting the greater influence of atmospheric variability on observations in the latter cases.

The increased self-consistency of ARC AATSR algorithms can be attributed to the improved forward modelling reported in Embury et al. (2012a). However, there is still some systematic discrepancy of order 0.1 K between two-channel and three-channel retrievals for middling levels of atmospheric TCWV. What level of forward modelling error does this imply? Merchant and Le Borgne (2004) pointed out that forward modelling errors tend to get amplified by a factor of the order of the sum of absolute weightings, which usually exceeds 1. Here, this amplification (assuming correlated errors between nadir and forward views of a given channel) ranges from about 2 (for D3 and N3) to 6 or more (for N2 at highest TCWV band). Agreement to ~0.1 K therefore implies that the forward modelling process is accurate to <0.1 K (with one caveat discussed below, related to the AATSR 12 µm channel).

Nonetheless, the ARC target is that systematic biases in SST be <0.1 K (Merchant et al., 2008c), and this is not fully met. We consider how the relative bias between two-channel and three-channel estimates can be reduced in the next section.

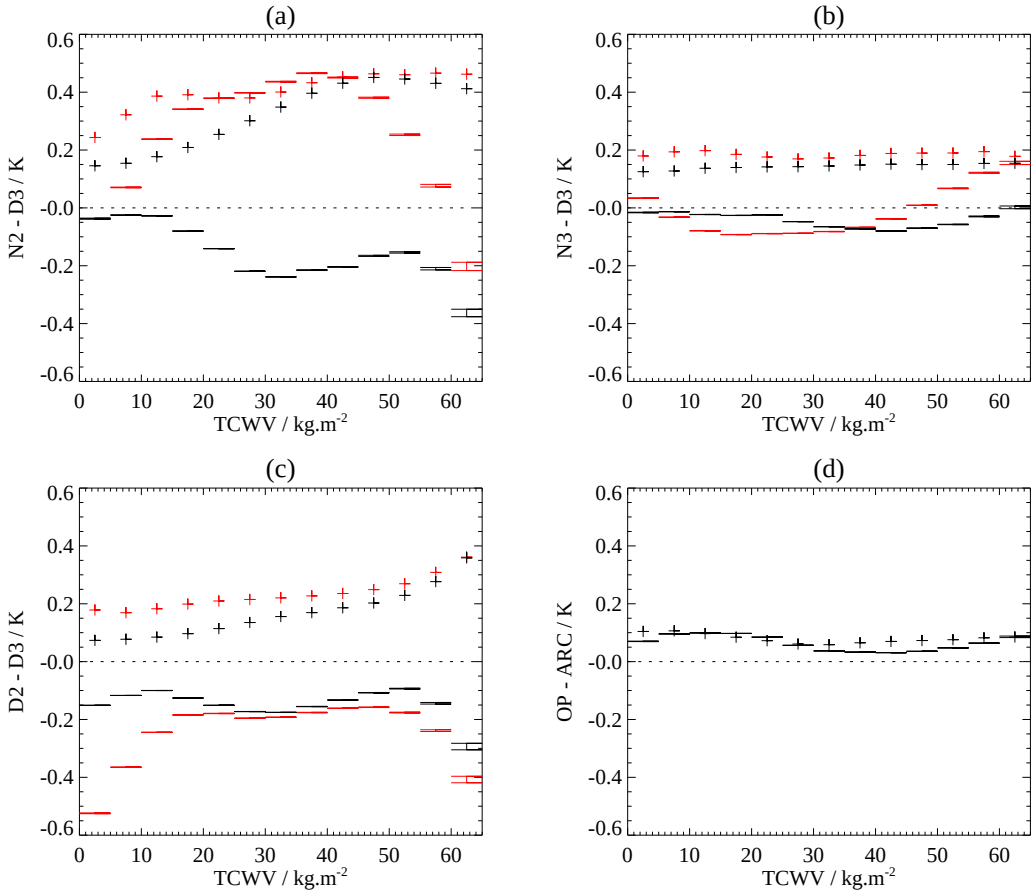

Figure 3.10: Inter-algorithm differences of operational (red) and new (black) retrieval coefficients for (a) N2-D3, (b) N3-D3, and (c) D2-D3. Also shown (d) OP-ARC difference for D3 retrieval. Error bars show mean ± standard error; crosses indicate standard deviation.

3.4.2 Mutual bias correction

Given that the relative bias between the ARC two-channel and three-channel SSTs has a modest but clear dependence on TCWV (Fig. 3.10), the different types of SST can be brought into better consistency by mutual bias correction of the offset coefficient, a_0, for each band of TCWV. Since this is done without reference to any in situ SSTs, independence of the satellite SSTs is preserved. However, which retrieval should act as the reference to which the others are corrected? The most convincing choice among the traditional algorithms is D3, for the following four reasons.

First, the D3 algorithm depends strongly on the 3.7 µm channel. The extremely steep dependence of radiance, L, at 3.7 µm on temperature in Planck's function is then an advantage. For clear-sky scene temperatures, the dependence is $L \sim T^n$ where n is in the range 13 to 15. This makes BTs at 3.7 µm easier to model reliably in some respects, because a given fractional error in L corresponds to a smaller error in BT than at 11 or 12 µm. Thus, uncertainty of a given magnitude in factors such as emissivity, reflected sky radiance, or absorption by atmospheric aerosols leads to relatively smaller errors in BT for the 3.7 µm channel.

The second point relates to the fact that the magnitude of the retrieval coefficients for three-channel algorithms tends to be several times smaller than coefficients in two-channel algorithms. This means there is less amplification of forward model errors when the 3.7 µm channel is part of the retrieval.

Thirdly, in simulation, the D3 algorithm has minimal geographical or functional biases (see subsection 3.3.3).

Finally, we prefer to use the D3 rather than N3 as the reference, because of the former's greater tolerance to the presence of tropospheric aerosols in any of the matches.

However, at this point we must consider a complication that requires a solution other than using D3 as a reference SST retrieval for the other algorithms. There is a significant calibration uncertainty in the 12 µm channel of AATSR. The details of this calibration uncertainty cannot be rehearsed in this paper. A fixed adjustment of −0.2 K to AATSR 12 µm BTs gives an approximate correction to the calibration (G Corlett, personal communication). This correction has been applied up front within the ARC processing. Work continues to develop a better understanding of the problem with this channel, but has not reached a conclusion (D. Smith, personal communication). Although the simple correction is apparently quite effective, there remains a greater uncertainty in the calibration of this channel. It is not satisfactory to use this AATSR channel in an algorithm which acts as a reference for the others. Instead, we use the D2* algorithm, which uses dual-view and only the 3.7 and 11 µm channels. (The * is to distinguish that the 3.7 and 11 µm channels are used rather than the 11 and 12 µm channels of the D2 algorithm.)

In simulation, D2* gives SST estimates that are practically indistinguishable from those obtained by D3: simulated residuals typically differ between D2* and D3 SSTs by <0.01 K apart from very dry atmospheres where differences can be up to 0.02 K as shown in Fig. 3.11. The four arguments listed

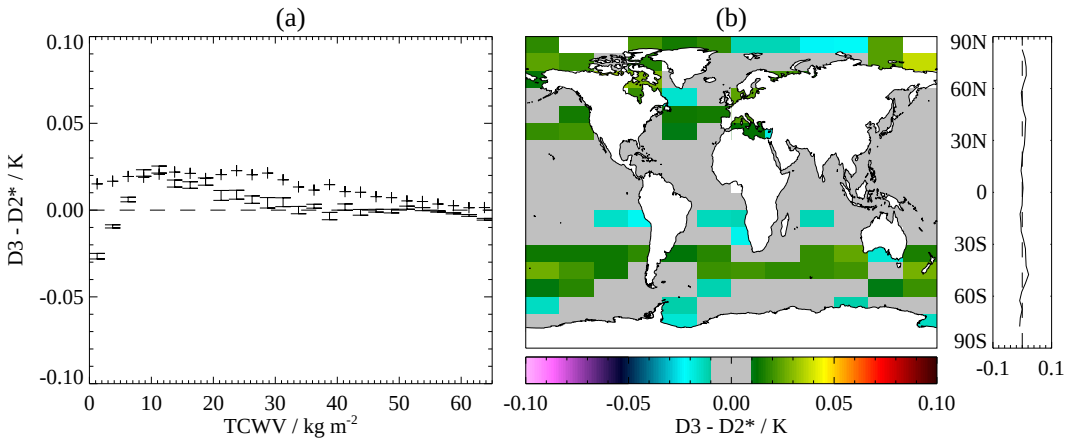

Figure 3.11: Simulated difference between D3 and D2* retrieval coefficients. Note full scale is only ±0.1 K.

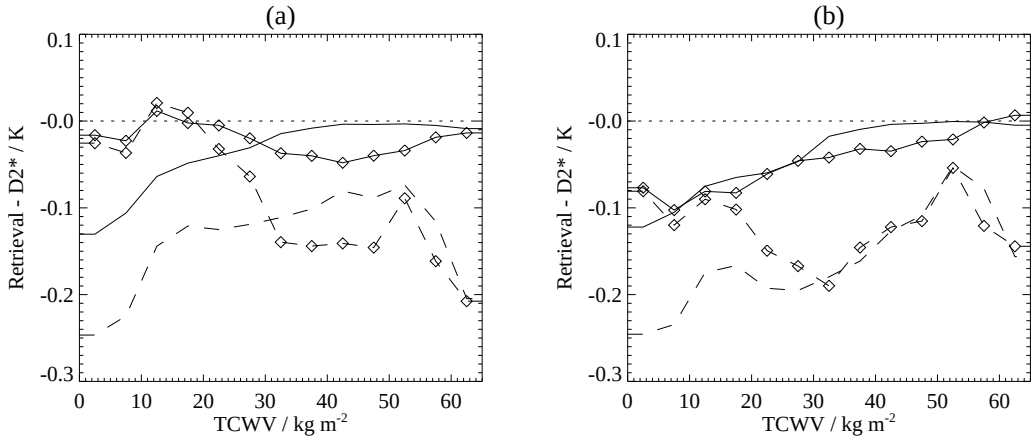

Figure 3.12: Offset adjustments to bring D3, D2, N3, and N2 coefficients into consistency with D2* coefficients for (a) ATSR-2 and (b) AATSR. Solid line: D3; dashed line: D2; solid with diamonds: N3; dashed with diamonds: N2.

above apply equally to D2* as to D3. Despite the calibration uncertainty being an AATSR issue, we use the D2* estimator as a reference for both ATSR-2 and AATSR, in order to maximize consistency of approach between the sensors.

The summary of how the final ARC coefficients are obtained is therefore as follows. N2, N3, D2, D3 and D2* coefficients for each sensor are derived from the simulations of BTs using Eq. 3.6. The coefficients are banded in time, nadir zenith angle, forward zenith angle and TCWV, as discussed above. A large global sample of observed clear-sky BTs is obtained; we use the set of matches to drifting buoys for convenience, but do not use the in situ SSTs in the coefficient definition. The retrieval coefficients are applied to the sample of clear-sky BTs and the bias between the four standard retrievals and the D2* retrieval is calculated for each N2, N3, D2, and D3 TCWV band. These biases are then subtracted from the a_0 offset of the coefficients in order to give retrievals that are consistent with D2* and are also consistent with each other.

Figure 3.12 shows the corrections found for ATSR-2 and AATSR. Although the D2* and D3 retrievals appear almost identical in simulations, there is a noticeable adjustment required in practice, with a ~0.1 K difference between D2* and D3 for low TCWV. As the same difference is seen for both ATSR-2 and AATSR it is unlikely that this is related to the 12 µm calibration problem, but instead appears to be another manifestation of the problem causing the 0.1 K bias between the N3/D3 and N2/D2 retrievals. For highly humid atmospheres the transmission in the 11/12 µm is very low and the matrix inversion assigns low weights to those channels in the D3 retrieval, giving coefficients close to D2*. For dry atmospheres the 11/12 µm transmission is higher so larger weights are assigned to these channels, i.e. the D3 coefficients become "closer" to D2.

3.5 Simulated characteristics of retrievals

Merchant et al. (2009) proposed the following characteristics as desirable for new SST retrieval schemes.

1. On scales upwards of ~1000 km, bias should be small (say, <0.1 K)
2. Sensitivity to atmospheric water vapour variability should be negligible (close to zero)
3. Sensitivity to true changes in skin SST should be close to unity (i.e., perfect sensitivity)

The authors then showed that these characteristics could be evaluated for empirically based AVHRR SST coefficients using a blend of in situ data and radiative transfer modelling. Here, we have already addressed the first characteristic in simulation with Fig. 3.8. These show the minimum magnitude of bias we can expect from the intrinsic properties of the retrieval scheme were the forward model perfect. The criteria for bias less than 0.1 K is met regionally by the TCWV-banded coefficient scheme in simulation. The initial assessment against in situ data in the companion paper (Embury et al., 2012b) demonstrates that this criterion is achieved in practice for AATSR.

The second criterion is assessed using calculations of the rate of change of BT with respect to TCWV ($\frac{\partial \mathbf{y}}{\partial w}$). The sensitivity of the retrieved SST to TCWV is then given by:

$$\frac{\partial \hat{x}}{\partial w} = \mathbf{a}^{\mathrm{T}} \frac{\partial \mathbf{y}}{\partial w} \tag{3.8}$$

Figure 3.13 shows this sensitivity as a regional map of bias, presented as the change induced by a 10% increase in water vapour at all levels of the atmosphere. The retrieval most sensitive to fluctuations in water vapour is, unsurprisingly, N2 where in tropical regions a 10% increase in water vapour can cause a $\sim\frac{1}{3}$ K decrease in the retrieved SST (the maximum sensitivity for any cell is −0.42 K). Outside the tropics this sensitivity rapidly decreases, and polewards of 45° north or south most cells have a sensitivity <0.01 K to a 10% change in water vapour. There are a few cells with positive sensitivity, i.e. the retrieved SST increases by more than 0.01 K for a 10% increase in w.

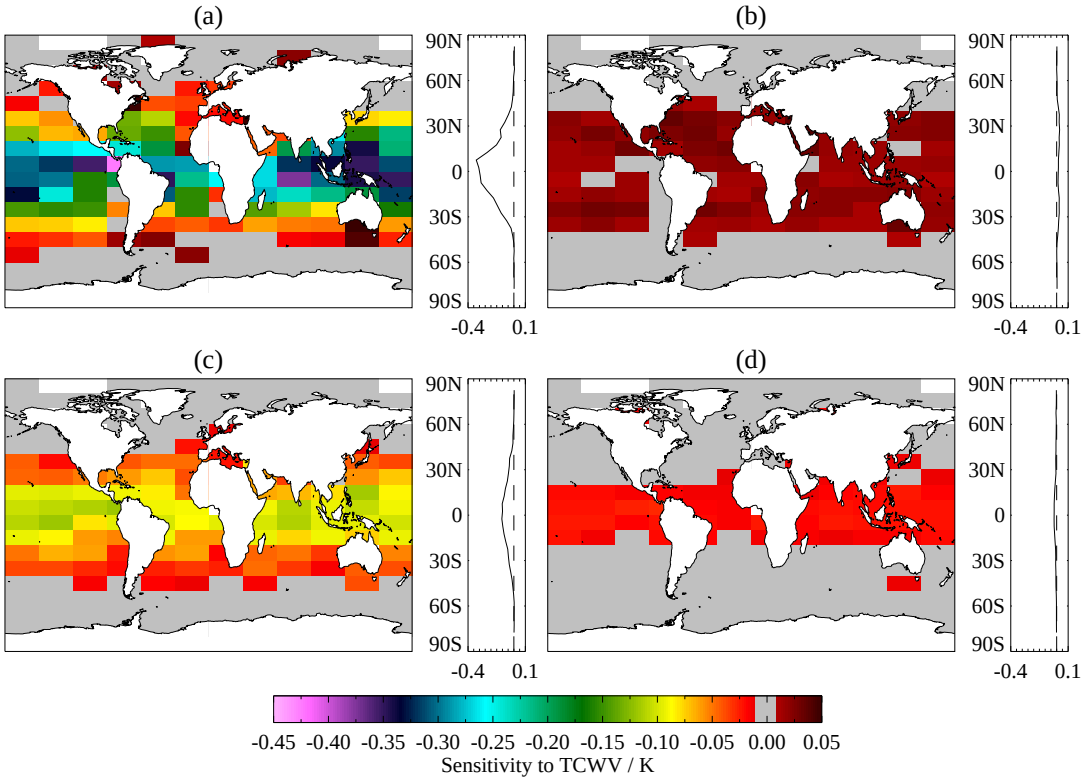

Figure 3.13: Sensitivity of retrieval algorithm to a 10% change in TCWV for (a) N2, (b) N3, (c) D2, and (d) D3 coefficients. Cells are 30° in longitude by 10° in latitude, with values <0.01 K considered negligible, in grey.

The next most affected retrieval is D2 (Fig. 3.13c), which follows a similar pattern to N2. In the tropics we generally see a ~0.1 K decrease for 10% increase in w; this falls off through the mid-latitudes and is essentially negligible polewards of 45°. Unlike the N2 retrieval there are no cases of positive sensitivity to w greater than 0.01 K.

Figure 3.13b shows the N3 retrieval, which behaves somewhat differently to the other retrievals. Here a 10% increase in w leads to an overestimation of the SST by ~0.015 between 45°S and 45°N (maximum sensitivity of 0.02 K). This is due to the use of the 3.7 µm channel which is far less sensitive to water vapour than the 11 and 12 µm channels. The D3 retrieval (Fig. 3.13d), however, does not show positive bias as it has the added information from the forward view. Sensitivity throughout the mid-latitudes and polar regions is <0.01 K. In tropical regions a 10% increase in w causes only a ~0.015 K decrease in retrieved SST.

In addition to changes to actual TCWV, the retrievals are also sensitive to changes in the prior TCWV which is used to select the appropriate TCWV band. We find that using water vapour bands of width 20 kg m^{-2} ensures that errors in prior TCWV have less impact than fluctuations in actual water vapour. The D3 is insensitive to errors in prior TCWV of 10% with change in retrieved SST always <0.01 K. N3 is more affected by errors in prior TCWV with a ~0.02 K impact between 45°S and 45°N. N2 and D2 are both approximately half as sensitive to errors in prior TCWV than to actual TCWV. Using narrower water vapour bands will lead to improved systematic biases, but unacceptably high sensitivity to both actual TCWV fluctuations and errors in prior TCWV (not shown).

The final criterion is assessed using the rate of change of BT with respect to SST. The sensitivity of the retrieved SST to a change in actual SST is:

$$\frac{\partial \hat{x}}{\partial x} = \mathbf{a}^{\mathrm{T}} \frac{\partial \mathbf{y}}{\partial x} \tag{3.9}$$

In the ideal case $\frac{\partial \hat{x}}{\partial x} = 1$ – that is a 1 K change in SST will result in as a 1 K change in the retrieved SST. Figure 3.14 shows the regional sensitivity to change in actual SST for the four retrieval types. N3 and D3 retrievals both have near-perfect sensitivity i.e. 1±0.01 K/K. D2, however, is slightly over sensitive to changes in actual SST; in the tropics a 1 K increase in SST will be detected as a slightly larger ~1.02 K increase in retrieved SST, with a few equatorial cells having sensitivities as high as ~1.04 K/K. The co-variability of SST and TCWV may mean that the over sensitivity to SST will be somewhat offset by the sensitivity to TCWV. The N2 retrieval is the most variable; mid-latitude and polar regions are over sensitive ~1.02 K/K, while tropical regions are under sensitive ~0.95 K/K, with a few regions as low as ~0.90 K/K.

The N2 algorithm uses the equivalent channels as the non-linear SST (NLSST) used for AVHRR in the Pathfinder project (Kilpatrick et al., 2001). Comparing the sensitivity of the N2 SST against the NLSST sensitivities reported in Merchant et al. (2009) we see similar zonal patterns but different regional details. The average NLSST response to TCWV is greater than the N2 response, ranging from (~) – 0.5 K for a 10% increase in TCWV in the tropics to slightly over 0.1 K in some mid-latitude regions.

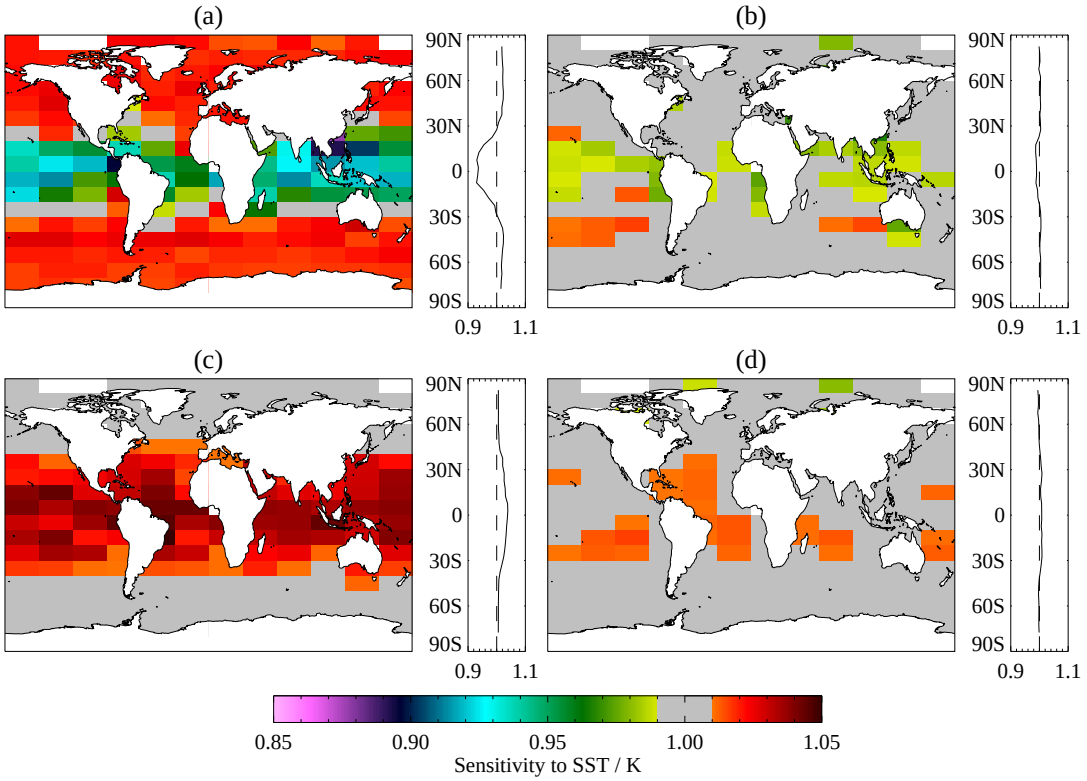

Figure 3.14: As Fig 3.13, but showing sensitivity to actual SST.

The average NLSST retrieval is also much less sensitive to change in true SST than is the case for N2, with the minimum sensitivity in tropical regions <0.5 K/K. The difference in SST sensitivity arises from the much wider AVHRR swath resulting in satellite zenith angles up to 55° compared with 22° for ATSR instruments.

Sensitivity to SST and insensitivity to TCWV are both importance characteristics of a SST retrieval scheme, the ideal being unit response to changes in true SST and zero response to TCWV fluctuations. In practice three channel retrievals will be closer to this ideal than the split-window due to the higher transmittance of the 3.7 µm channel. There are several consequences of non-ideal responses. For instance, non unit sensitivity to true SST changes means that temperature differences across ocean fronts can be underestimated, as could diurnal warming events. Sensitivity to changes in TCWV means that atmospheric variations may be misinterpreted as ocean variability. Furthermore, when sensitivity to true SST is not unity, it follows that the retrieval is dependent on prior SST information. This can be explicit, such as the "guess-SST" used by Pathfinder, or implicit – i.e. embedded in the retrieval coefficients, which incorporate prior information inherent in the training dataset.

3.6 Discussion

In this paper we present the definition of the new linear retrieval scheme to be used within the ARC project for estimating SST and its relation to earlier algorithms for ATSR and AVHRRs. As with the

operational ATSR retrievals, we base our coefficients on radiative transfer simulations in order to maintain independence from in situ data. The main refinements to the algorithm itself are: the use of the satellite viewing geometry at time of observation rather than a nominal geometry; banding the coefficients by TCWV rather than by latitude; and bringing the coefficients for different channel combinations into consistency by adjusting their offsets with reference to the new D2* algorithm.

A consequence of basing coefficients on radiative transfer simulations that are, inevitably, imperfect is retrieval bias (Merchant and Le Borgne, 2004). In general, this bias is different for different channel combinations. The dual-view capability of the ATSR instruments allows a large number of different channel combinations for retrievals. The four in general use are dual-view and nadir-only versions of the split-window and three-channel algorithms. The degree of consistency between the different retrievals provides an indication of the forward model accuracy, without the need of an external reference (i.e. in situ data). The operational AATSR N3 and D3 retrievals give SSTs consistent to better than 0.1 K across most regions. However, D2 SSTs are biased ~0.2 K cold on average relative to D3, while N2 is ~0.4 K warmer on average, varying between −0.1 and +0.7 K regionally. Using coefficients derived from new radiative transfer simulations (Embury et al., 2012a), prior to mutual bias correction, we find that the N2 and D2 are now also consistent to better than 0.1 K offset in the global mean, although N2 and D2 are ~0.1 K cold relative to the N3 and D3 SSTs. Only N2 shows large regional variations in relative bias compared to the other algorithms, of between −0.4 and +0.2 K compared to D3.

All linear retrievals are subject to complex systematic errors, due to the use of a linear algorithm to solve a non-linear problem (Merchant et al., 2006b). While such issues can be addressed through the use of an optimal estimation retrieval (Merchant et al., 2008b), this can introduce larger forward modelling errors because it requires the use of a fast parametrised forward model. Alternatively, independent information such as a prior SST or TCWV estimate can be added to the retrievals through the use of the first-guess as a term in the retrieval or by stratifying the coefficients. While the use of a first-guess SST is fairly common in non-linear retrievals, we find it more effective to use water vapour banded coefficients. Regional bias simulation experiments show the TCWV banded coefficient are significantly improved compared with the global coefficients, with three of the ATSR channel combinations (D2, N3, and D3) meeting the target of regional biases <0.1 K.

While the new radiative transfer simulations have led to improved inter-algorithm biases compared to the operational coefficients, the N2 and D2 SSTs remain ~0.1 K colder than the N3 and D3 SSTs. The three-channel retrievals are only used at night, while two channel retrievals must be used for daytime data. Thus, the ~0.1 K difference will introduce a spurious day-night difference. Therefore, it is necessary to bring the four retrievals into self-consistency by adjustment to a reference retrieval (without use of in situ observations). Because of the uncertain calibration of the 12 µm channel on AATSR, we have introduced a dual-view retrieval D2* which uses only the 3.7 and 11 µm channels to act as this reference. After this step, all algorithms are unbiased with respect to each other. Regional differences are <0.1 K for N3 and D2 in comparison with D3. This is mostly true for N2 also, except for

deviations of magnitude 0.2 K in the most extreme tropical conditions.

In addition to regional bias, we evaluate the coefficients using the criteria from Merchant et al. (2009) – i.e. the retrievals should, in simulation, have unit sensitivity to changes in SST and zero sensitivity to changes in TCWV. Both these properties are important when measuring SST for climate. If a retrieval has non-unit response to SST changes, any observed temperature trends will be biased by prior error. Non-zero sensitivity to TCWV means long term trends in atmospheric state may influence any observed SST trend. The three channel N3 and D3 SSTs both show excellent characteristics with sensitivities generally within 0.01 of the ideal. The D2 SSTs are still of good quality, but may overestimate SST variability in the tropics by 2–4%, and are expected to have a ~0.1 K response to a 10% change atmospheric water vapour in the tropics.

New linear retrieval coefficients have been designed for the ATSR Reprocessing for Climate project, with the dual-view coefficients meeting, in simulation, stringent requirements for climate. The nadir-only N3 also meet the requirements for climate, in the absence of stratospheric aerosol. The N2 retrieval does not meet the regional bias requirements and has larger sensitivity to atmospheric water vapour making it less suitable for climate applications; it is still of excellent quality for other uses. Furthermore the improved consistency of all four retrievals, including N2, in inter-algorithm comparisons validates the underlying radiative transfer simulations. The companion paper Embury et al. (2012b) compares the new ARC SSTs against in situ data.

The developments reported here are applicable also to ATSR-1. However, ATSR-1 will require additional measures to account for the effects on spectral response and calibration of a change in detector temperatures (from ~80 K early in the mission to ~110 K towards the end). Thus, the scheme presented here will need to be extended, and this will be reported in a separate paper.

In the future, the Sea and Land Surface Temperature Radiometer (SLSTR) will be the dual-view sensor most comparable to the ATSRs. SLSTRs are planned to be launched by the European Space Agency on the Sentinel-3 platforms from 2013. All the elements of the coefficient scheme defined in this paper are transferable to SLSTR. Current plans are that SLSTR operational coefficients at launch will implement TCWV banding and account for dual-view zenith angles on the same basis as in this paper. Adjustment for consistency to a reference algorithm is only possible after launch, once in-flight observations are available.

Chapter 4

Initial validation, accounting for skin and diurnal variability effects

Owen Embury, Christopher J. Merchant, Gary K. Corlett

Abstract

An initial validation of the Along Track Scanning Radiometer (ATSR) Reprocessing for Climate (ARC) retrievals of sea surface temperature (SST) is presented. ATSR-2 and Advanced ATSR (AATSR) SST estimates are compared to drifting buoy and moored buoy observations over the period 1995 to 2008. The primary ATSR estimates are of skin SST, whereas buoys measure SST below the surface. Adjustment is therefore made for the skin effect, for diurnal stratification and for differences in buoy-satellite observation time. With such adjustments, satellite-in situ differences are consistent between day and night within ~0.01 K. Satellite-in situ differences are correlated with differences in observation time, because of the diurnal warming and cooling of the ocean. The data are used to verify the average behaviour of physical and empirical models of the warming/cooling rates.

Systematic differences between adjusted AATSR and in-situ SSTs against latitude, total column water vapour (TCWV), and wind speed are less than 0.1 K, for all except the most extreme cases (TCWV<5 kg m^{-2}, TCWV>60 kg m^{-2}). For all types of retrieval except the nadir-only two-channel (N2), regional biases are less than 0.1 K for 80% of the ocean. Global comparison against drifting buoys shows night time dual-view two-channel (D2) SSTs are warm by 0.06±0.23 K and dual-view three-channel (D3) SSTs are warm by 0.06±0.21 K (day-time D2: 0.07±0.23 K). Nadir-only results are N2: 0.03±0.33 K and N3: 0.03±0.19 K showing the improved inter-algorithm consistency to ~0.02 K. This represents a marked improvement from the existing operational retrieval algorithms for which inter-algorithm inconsistency is >0.5 K. Comparison against tropical moored buoys, which are more accurate than drifting buoys, gives lower error estimates (N3: 0.02±0.13 K, D2: 0.03±0.18 K). Compara-

ble results are obtained for ATSR-2, except that the ATSR-2 SSTs are around 0.1 K warm compared to AATSR.

4.1 Introduction

A new sea surface temperature (SST) retrieval has been developed for the Along Track Scanning Radiometer (ATSR) instruments, with a view to developing an independent SST time-series for climate applications (Merchant et al., 2008c). Accurate retrieval of SST was the primary motivation for the ATSR missions (for a review and references, see Llewellyn-Jones et al., [this issue]). The instruments are well calibrated, and use along-track scanning to give both a nadir view of the ocean surface and a forward view at a satellite zenith angle of around 55 degrees. The dual view capability is particularly important in ensuring continuity of accuracy in anomalous conditions, such as when stratospheric volcanic aerosol is present. ATSR SST retrieval coefficients have developed significantly (Merchant and Harris, 1999) since the work of Závody et al. (1995). The results presented here arise from the project ATSR Reprocessing for Climate (ARC) (Merchant et al., 2008c). Companion papers in this issue have detailed the ARC's radiative transfer simulations (Embury et al., 2012a) and improved design of retrieval coefficients (Embury and Merchant, 2012), which includes a refined treatment of the ATSR viewing geometry and the use of coefficients banded by total column water vapour to reduce biases with respect to atmospheric state (Barton, 1998; Merchant et al., 2006b). As a point of comparison, the ARC results are at points compared with results from techniques used in creating the ATSR SSTs in the ATSR multi-mission archive version 2.0.

This paper presents an initial comparison of the new satellite-derived SST estimates obtained for ATSR-2 and Advanced ATSR (AATSR) with independent measurements.

There are two novel aspects to the work with wider ramifications, as follows.

Firstly, in section 4.2, we discuss how the purpose for which a validation is undertaken affects the quality control to be applied and the statistics to be reported. Here, we apply minimal quality control of in situ observations and report robust statistics, for reasons given.

Secondly, as described in section 4.3, the satellite-derived estimates validated here are of two types. The primary observations are of skin SSTs (estimates from the satellite observations of the infrared radiometric temperature of the sea surface at the time of observation). In a further step, we derive time-adjusted depth SSTs (estimates adjusted for the ocean thermal skin effect, near-surface stratification and differences in observation time). The latter are intended to be more directly comparable to the in situ data used for validation. It is shown how adjustment for differences in time of observation can be modelled from a sufficiently large data set itself, and that this reduces systematic effects in the residuals between satellite and in situ SSTs (section 4.4).

Nevertheless, the main purpose of the paper is the initial validation of new SSTs obtained within the ATSR Reprocessing for Climate project (ARC) (Merchant et al., 2008c). The validation is "initial" in the sense that all results presented here are derived before homogenisation (blending) of the multi-

satellite SST time series. Ideally, the SSTs obtained from ATSR-2 and AATSR would be so consistent that no homogenisation would be necessary. Practically, the aim of the careful work done on retrieval development is to minimize any homogenisation required, since this increases the confidence we can have in the stability of the homogenised record that will ultimately be created. We find here that the global and regional retrieval biases in the new satellite SSTs are low, mostly <0.1 K. The discrepancy between ATSR-2 and AATSR is also found to be <0.1 K. Although small, this discrepancy is not negligible. Homogenisation will therefore be a crucial step in developing a record appropriate for climate, and will be reported in a subsequent article.

4.2 Approach to validation

4.2.1 Statistical approach

It is important to be clear about the purpose of the SST validation reported here. Our objective is to assess the accuracy of the new SST retrieval scheme defined in Embury and Merchant (2012). This is a narrower focus than validation of a new SST product, since the quality of an SST product is determined by the whole satellite-processing system, which includes determination of which pixels are valid for SST retrieval (by cloud and sea ice detection, checks on validity of radiances, etc). Cloud detection failures in particular are known to introduce occasional large errors in retrieved SST that significantly increase the standard deviation of satellite SSTs in comparisons with in situ SSTs. In this paper, we make the assumption that the incidence of cloud detection failure is low (this is shown to be reasonable in section 4.2.2).

Outliers can also arise from in situ errors. Operational systems for flagging erroneous in situ data typically offer a basic level of quality control (QC), and it is usual for additional QC to be applied (e.g., O'Carroll et al., 2008). However, any QC to remove doubtful data inevitably involves a significant element of judgement: the harder one looks, the more cases one can find justifications for rejecting. There is therefore an arbitrary element that makes reported statistics difficult to compare between validation studies: to some degree, the validation statistics validate the judgements made regarding QC and not just the SST retrieval method. In this work, the in situ observations are subject to only minimal additional QC (e.g., in situ location is confirmed to be ocean, temperature is not below freezing).

As our purpose is to assess the new SST retrieval scheme, we are not concerned with characterizing the distribution of outliers that arise for the reasons above. Therefore we use outlier tolerant statistics, also known as robust statistics (Huber, 1981), to assess the retrieval results. No further attempt is made to catch and eliminate failures of cloud detection or outliers in the in situ data during the statistical analysis.

To describe the central value of a distribution we use the median. For a Gaussian distribution or any symmetric distribution, this would equal the mean. To describe the spread of a distribution, we

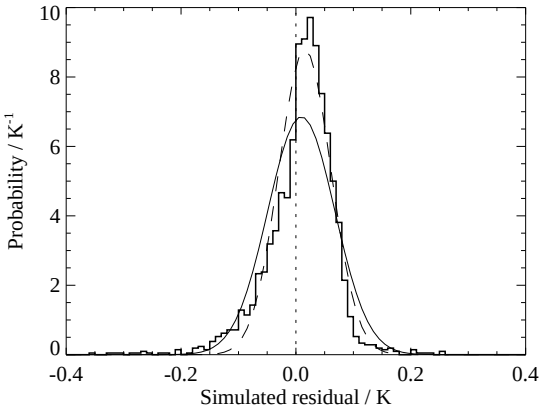

Figure 4.1: Probability distribution of simulated D2 retrieval error. Thin solid line shows Gaussian distribution using standard statistics, the standard deviation (SD) being 0.058 K; thin dashed line shows Gaussian distribution using robust statistics, the robust SD being 0.045 K. There are 2100 simulated cases, as described in Embury and Merchant (2012), and those shown are for the AATSR instrument. Results for other ATSRs are similar.

use a "robust standard deviation" or RSD. The RSD used is 1.48 times the median absolute deviation from the median (MAD). For a Gaussian distribution, the RSD and conventional standard deviation are equal. There are many other ways to reach comparable statistics, such as finding the best fit Gaussian, scaling the inter-quartile range, or applying a 3-sigma filter and recalculating conventional statistics on a second pass. However, the median and RSD are easy to calculate and have a simple analogy to their conventional counterparts. Both measures are highly robust meaning that a significant minority of data points can be replaced with outliers of infinity without the calculated statistics also going to infinity. In the particular case of SST validation, these robust statistics are not overly sensitive to cloud detection or gross in situ errors, and therefore isolate the performance of the SST retrieval itself.

It could also be argued that the retrieval itself is a source of outliers arising from extreme atmospheric conditions (e.g., Minnett, 1986). This is true, but we cannot identify in a validation data set which outliers do arise in this way, rather than from cloud detection or in situ gross errors. The best information we have is from radiative transfer simulations for a comprehensive range of atmospheric states (Embury et al., 2012a). Figure 4.1 shows the distribution of simulated retrieval error for dual-view 11 and 12 µm retrieval, and shows that while the distribution is slightly skewed there are relatively few outliers and 96.5% of data points lie within 3-RSDs of the median. The outliers are mostly associated with high amounts of atmospheric water vapour. Table 4.1 shows various statistical measures of the simulated retrieval errors. The statistics are given for all four types of retrieval as discussed in Embury and Merchant (2012): N2 and N3, these being single-view retrievals using 11 and 12 µm and 3.7, 11 and 12 µm channel sets respectively; and D2 and D3, the corresponding dual-view combinations. In all cases the mean and median agree to better than 0.01 K, and only the nadir-only two-channel retrieval (N2) has markedly different SD and RSD.

	Min	LQR	Median	RSD	Mean	SD	UQR	Max
N2	-1.127	-0.099	0.015	0.157	0.010	0.284	0.116	1.462
N3	-0.239	-0.049	-0.005	0.067	0.003	0.080	0.044	0.584
D2	-0.359	-0.018	0.016	0.045	0.009	0.058	0.045	0.256
D3	-0.146	-0.024	0.002	0.039	0.003	0.039	0.028	0.187

Table 4.1: Summary statistics of simulated retrieval error in the absence of instrumental noise for retrievals using four channel combinations, N2, N3, D2 and D3 (see main text for definitions). "Min" and "Max" are the minimum and maximum simulated retrieval error; "LQR" and "UQR" are the lower and upper quartiles; "RSD" is the robust standard deviation (as defined in main text); "SD" is the standard deviation. All quantities are in kelvin.

4.2.2 Match-up data

In situ data

For the purpose of this initial validation, we created a match-up database (MDB) that contains both in situ SST measurements and the collocated ATSR observations. The in situ SST observations are provided by the Met Office Hadley Centre in the form of ASCII files using one decimal place accuracy, i.e. in situ latitude and longitude are reported to 0.1 degree precision and SST measurements to 0.1 K, and are comprised of data from the International Comprehensive Ocean-Atmosphere Dataset (ICOADS, Worley et al., 2005) from the start of the ATSR-1 mission to the end of 1997, and National Centers for Environmental Prediction (NCEP) near-real-time (NRT) marine reports (see http://icoads.noaa.gov/nrt.html) from the start of 1998 to present. The majority of data are received through the Global Telecommunication System (GTS) which includes data from three main sources: drifting buoys, moored buoys, and ships. Additional QC information using the method of Rayner et al. (2006) is provided with the in situ data, from which we exclude obviously bad data before matching with the satellite data. Checks included in the QC flags are: invalid or missing data; blacklisted data sources; inland and lake buoys; temperatures below −1.8°C; temperatures > 8°C different from climatology; and a buddy check (in which nearby buoys are compared).

Each of the three in situ data types has particular characteristics. Drifting buoys have the most complete spatial coverage, although it is far from uniform, and they measure the sea temperature at around 20 cm depth. The uncertainty in drifting buoy observations was estimated to be ~0.2 K by O'Carroll et al. (2008), and part of the uncertainty arises from reporting of drifting buoy temperatures with only 0.1 K resolution (i.e., rounded to one decimal place). During the AATSR mission drifting buoys formed the majority of the in situ dataset. Moored buoys can be roughly separated into two distinct types, the Global Tropical Moored Buoy Array (GTMBA) and coastal moorings. The GTMBA includes data from the TAO/TRITON (Tropical Atmosphere Ocean / Triangle Trans Ocean Buoy Network; McPhaden, 1995), PIRATA (Prediction and Research Moored Array in the Atlantic; Bourlès et al., 2008), and RAMA (Research Moored Array for African-Asian-Australian Monsoon Analysis and Prediction McPhaden et al., 2009) arrays, and all use ATLAS (Autonomous Temperature Line Acquisition System) moorings. These provide the most accurate measurements in the ATSR MDB: the

instrument accuracy is better than the 0.1 K resolution reported in GTS records, but the number of match ups is low, representing <3% of the MDB. ATLAS moorings measure SSTs at 1 m depth. Ouside of the GTMBA, most other moored buoys are located in North American and European coastal waters. While coastal moored buoys do record more information than drifters, their designs are more varied hence temperature measurements are from a wider range of depths. Also, the coastal location is particuarly challenging for SST retrieval from space, and so the coastal moored buoys are excluded from this initial ARC SST validation. Finally, there are measurements from Voluntary Observing Ships (VOS). These represented the majority of the in situ record at the start of the ATSR1 mission (1991); however, the share was rapidly reduced both due to the reduction in total number of VOS reports and the increase in buoy data. VOS measurements use a range of different sensors – hull contact, engine intake etc. – resulting in significantly poorer precision than buoy data (e.g. Emery et al., 2001), and therefore are also excluded from the initial ARC SST validation.

Satellite data

Matching the in situ data with the satellite imagery is done as follows. Firstly the in situ data are filtered to select only reports within three hours of a given ATSR data-file (corresponding to a single orbit); these are then collocated to the nearest ATSR pixel (whether clear or cloudy) and all matches with separation >1 km (i.e. outside the ATSR swath) are rejected. Secondly, matches to duplicate satellite data are detected and removed. These are cases where a single buoy has reported multiple observations, within the ±3 hour window, that are matched to the same ATSR observation. The match with the smallest time separation is retained and the remaining are ignored. Matches where a single buoy observation is matched against more than one ATSR orbit can occur at higher latitudes where the same location is observed on consecutive overpasses. Such matches are considered valid and are retained.

While matching against the pixel level product in this way should ensure that all satellite-in situ distances are less than 1 km, in practice it is limited by the geolocation accuracy of the two data types. As the in situ locations are reported to only 0.1° resolution, spatial separations of up to ~7 km may exist. ATSR SSTs are extracted for a 5 × 5 pixel window surrounding the nominal in situ position and the average SST is calculated using all the clear-sky pixels. This pixel window reduces the random SST uncertainty from sensor noise by a factor of up to $5 = \sqrt{25}$.

Additionally, various numeric weather prediction (NWP) data used in the ARC processing system, such as surface wind speed and total column water vapour (TCWV), are extracted along with the satellite retrievals.

There are two consequences from the approach of matching to full-resolution data, rather than spatially averaged SST products as done in some previous AATSR validation studies (e.g., drifter validation in Corlett et al., 2006; O'Carroll et al., 2006). The first is that this allows the satellite-in situ spatial and temporal separation to be more accurately estimated and minimised. The second relates

to the fact that the operational ATSR spatially averaged products are generated by first averaging the brightness temperatures and retrieving the SST from those averages. The dual-view nature of the ATSR instruments means that different pixels can be cloudy in the two views. This can introduce unpredictable errors into the dual-view retrievals as the nadir and forward views are no longer guaranteed to correspond to the same source locations. We calculate SSTs from individual pixel brightness temperatures, ensuring the locations match as far as is possible between nadir and forward views.

The cloud screening algorithm used in the ARC processing system is based on the probabilistic Bayesian method described in Merchant et al. (2005), updated for day-time observations using methods in Mackie et al. (2010), and further developed for simultaneous exploitation of the two views of ATSR. A future article will give a full description and report on validation of the implementation for ARC of the Bayesian cloud detection algorithm; here we simply comment that, overall, it seems more effective than the threshold-based "SADIST" method (Závody et al., 2000) used with operational ATSR data. Table 4.2 compares the SD and RSD of satellite-in situ differences for the two cloud masks. The standard deviations show a substantial decrease when using the Bayesian cloud detection, equivalent to removing independent noise of magnitude >0.3 K. Even the outlier-tolerant RSDs show modest improvement (equivalent to removing ~0.12 K of independent noise), suggesting that the improvement in cloud detection is not limited to outliers. For a dual view SST, the global retrieval bias is not greatly affected by the cloud detection used, probably because cloud contamination can cause warm or cold biases depending on which view is affected by residual cloud. The ratio of the SD to RSD gives a measure of how much a particular SD is affected by outliers (with a ratio of 1 implying no outliers). The ratios are closer to unity for the Bayesian cloud detection than for SADIST. Even for Bayesian detection, however, the ratio is well above 1, reflecting the fact that cloud detection is not the only contribution to outliers. The tropical moored buoys matched with Bayesian-screened SSTs have SD equal to 0.24 K, with the RSD equal to 0.17 K, reflecting both the improved cloud detection and higher quality of the in situ data compared to drifters.

Match-up database summary

To assess the new SST retrieval scheme used within ARC, we require that the comparison data are accurate, consistent, and cover the widest possible range of observing conditions. Drifting buoys provide the widest coverage of all the in situ types, with reasonable consistency and accuracy, and are used throughout this paper, except where otherwise stated. The GTMBA data are more accurate and consistent, but the low spatial coverage and number of matches make them unsuitable for fully characterising global SST. Nevertheless, because of their good accuracy they are also useful for comparison against the ATSR data. All comparisons from this point on are based on ARC cloud screening, irrespective of whether the SST retrieval uses operational or ARC coefficients. Because operational SSTs were derived only for image pixels flagged clear by the operational cloud flagging,

	Drifters			GTMBA		
	N	SD	RSD	N	SD	RSD
Day						
SADIST	89831	0.565	0.236	5256	0.370	0.220
Bayesian	120916	0.458	0.235	6837	0.324	0.218
Night						
SADIST	100826	0.602	0.257	3571	0.400	0.220
Bayesian	96857	0.452	0.227	3081	0.242	0.174

Table 4.2: Impact of different cloud detection on validation results, for AATSR between July 2002 and December 2007. Statistics are shown in kelvin for residuals between satellite and in situ observations when using cloud-flags from the operational imagery ("SADIST") and from ARC (Bayesian) cloud screening algorithms. SD represents the standard deviation of satellite-in situ differences. RSD is an outlier-tolerant robust estimate of standard deviation. The satellite SSTs are dual-view two-channel skin SSTs (D2 SSTs) based on the same ARC algorithm throughout, to isolate the effects of cloud screening alone. D2 SSTs are not the most accurate of the available types of retrieval, but are available day and night, allowing the impact of the different day and night cloud detection schemes to be compared.

	ATSR-2	AATSR
Start date	1995-06-01	2002-07-24
End date	2003-06-22	2007-12-24
Total drifters	397474	922464
Clear-sky drifters	83007	218826
Clear-sky drifters/year	9896	41945
Total GTMB	24557	31398
Clear-sky GTMB	7772	9918
Clear-sky GTMB/year	892	1872

Table 4.3: Drifting buoy matches for ATSR-2 and AATSR. "Clear-sky" indicates matches with at least one clear pixel in the 5 × 5 pixel array around the reported buoy location.

SSTs using operational coefficients have been derived by re-applying the operational coefficients to imagery cloud-screened by ARC methods.

Table 4.3 summarises the total number of matches found. Approximately 75% of all matches found corresponded to completely cloudy scenes – i.e. all 25 pixels in the 5 × 5 window were classified as cloud by the Bayesian cloud detection algorithm – and are not used for retrieval validation. Figure 4.2 shows the geographical distribution of the drifting buoy matches retained for validation for ATSR-2 and AATSR.

4.3 Satellite-Buoy differences

4.3.1 Skin SST and skin effect

Radiometers operating at infrared wavelengths, such as the ATSR instruments, are sensitive to radiation emitted from the layer between the air-sea interface and about 20 µm below the air-sea interface, depending on wavelength. In situ measurements of SST, however, are usually made at some distance below the surface, the exact depth dependent on the observing platform used. During night there is a temperature difference of approximately 0.2 K between the two measurements. This is a "skin effect",

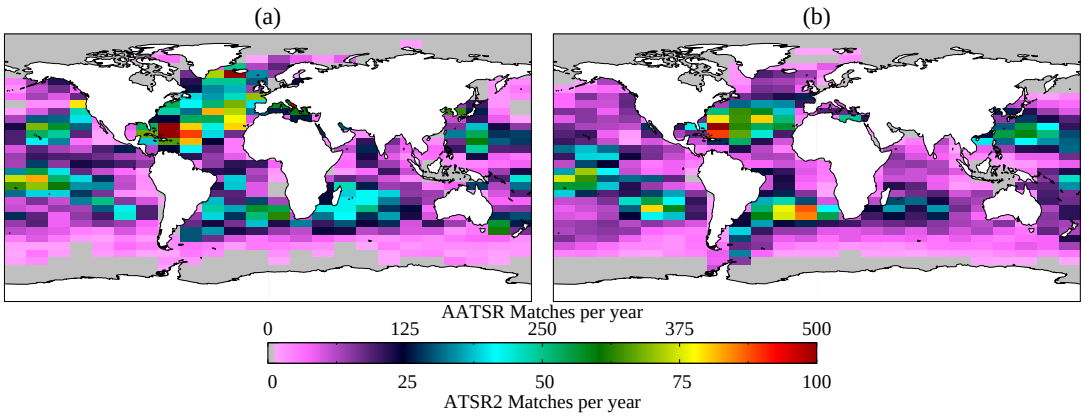

Figure 4.2: Global plots of clear-sky matches per year with drifting buoys for (a) ATSR-2 and (b) AATSR, per cell of 15° in longitude by 5° in latitude. The rate of matches can be compared in terms of matches per unit area by scaling the plotted value at a given location by the arc-cosine of its latitude.

reflecting the large thermal gradient required to maintain a flux of heat through the near-surface layer in which molecular rather than turbulent heat transfer processes dominate (e.g., Saunders, 1967). The skin effect exists also during the day, but may be modified by absorption within the ocean skin layer of near-infrared solar insolation. Under conditions of sufficiently low wind-driven mixing, solar heating will also cause thermal stratification of the upper ocean (discussed in the next section) further altering the difference between what is observed by satellite and measured by the in situ instrument. The Met Office has provided estimates of the skin effect for the ARC project using the Fairall et al. (1996) method forced by ECMWF-Interim data (Simmons et al., 2007). This accounts for both wind speed and net heat flux (full details of the implementation used are given in Horrocks et al., 2003a). The Fairall model includes two tuneable parameters (λ_0 and A), which have been set to $\lambda_0 = 4.5$ and $A = 0.2$ for this project, based on a previous study as discussed later in this section.

Figure 4.3 shows the difference between the AATSR retrieved SST_{skin} and SST_{depth} as measured by drifting buoys along with the predicted skin effect from the Fairall model output. For comparison we also show the empirical parameterization of night-time skin effect in terms of wind speed (u) from Donlon et al. (2002), namely:

$$SST_{subskin} - SST_{skin} = 0.14 + 0.30 \exp(-u/3.7) \tag{4.1}$$

In the nighttime case, other than a warm offset of ~0.05 K, the observed satellite-drifter differences follow the expected variation with wind speed predicted by the Donlon curve. The offset may represent an overall skin SST retrieval bias, a bias in the Donlon parameterisation, or a combination of these factors. The variations with wind speed of the Donlon estimate and the night time data are similar, in that there is little variation above 6 or 7 m s^{-1}.

Applying the Fairall method at night yields the shaded (\\) band. The band shows the median plus and minus one RSD of the result of the Fairall calculation; there is a range for a given wind speed

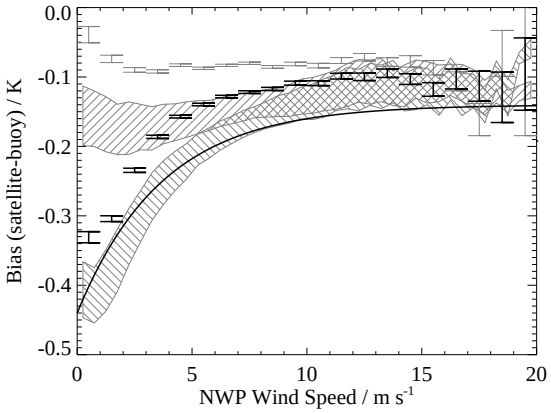

Figure 4.3: AATSR retrieved SST_{skin} minus drifting buoy SST as a function of wind speed. Error bars show the range of median difference plus and minus one standard error for (black) night-time and (grey) day-time matches. Solid line shows empirical parameterization of skin effect from Donlon et al. (2002). The shaded areas show the estimate using the Fairall et al. (1996) method, the vertical range being the median ± RSD for each bin of wind speed: (\\) night-time cases; (//) day time cases. Unlike the Donlon method, which is a function only of wind speed, Fairall's method attempts to account for heat flux through the skin layer and solar absorption within the skin.

because the Fairall method accounts for the heat flux variability through the skin layer. The Fairall skin-effect matches the observations at higher wind speeds (>10 m s^{-1}), cools with respect to the observed difference for winds in the range ~10 down to ~5 m s^{-1}, and is colder than but parallel to the observed difference at lower wind speeds.

During day-time, the absorption of insolation by the ocean skin leads to satellite-drifter differences which are almost independent of wind speed at ~0.1 K, except for a rise at wind speeds <3 m s^{-1}. The observed difference is a combination of the day-time skin effect and, particularly at wind speeds <3 m s^{-1}, thermal stratification (Murray et al., 2000) between the surface and the drifting buoy observation depth. The shaded (//) band is the day-time skin effect calculated using the Fairall method, which shows much less variation with wind speed than the nighttime data.

For the next step in comparison, we formulate an $SST_{subskin}$ estimate by using the Fairall model to adjust the SST_{skin} retrieval. Note that we could apply the adjustment to the drifting buoy SSTs to form a skin SST estimate to compare with the satellite SST_{skin}. The residual differences would be identical. The convention we adopt here means that all estimation steps (SST retrieval and adjustments) are kept on one side of the comparison and are validated against unadjusted in situ measurements.

Figure 4.4 shows the difference between the estimated $SST_{subskin}$ (satellite skin SST adjusted using the Fairall method) and drifter SSTs as a function of windspeed. The $SST_{subskin}$-drifter difference is <0.03 K for higher wind speeds (>10 m s^{-1}), with a warmer bias of up to 0.1 K for lower wind speeds. The day and night $SST_{subskin}$-drifter differences are now very consistent – for all wind speeds above ~3 m s^{-1} the day and night difference agree to ~0.02 K. A plausible interpretation is that the divergence of the day and night differences for winds <3 m s^{-1} is the consequence of thermal stratification within the ~20 cm of the surface (between the drifter thermistor depth and the surface). There is good

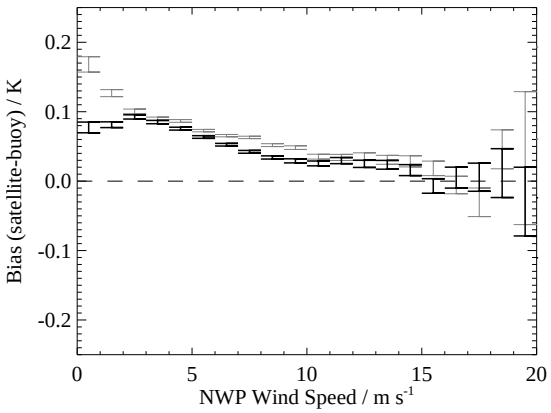

Figure 4.4: $SST_{subskin}$ minus drifting buoy SST as a function of wind speed. The $SST_{subskin}$ estimate is the AATSR-retrieved skin SST adjusted for the skin effect using the Fairall model (see main text). Error bars span the range of median difference plus and minus one standard error for (black) night-time and (grey) day-time matches.

consistency of the day and night curves at higher wind speeds (3 to 15 m s^{-1}) despite quite distinct day and night behaviour over this range in Fig. 4.3. This suggests that the Fairall formulation is an effective tool for estimating the different wind-dependence of the skin effect in day and night. The residual variation of bias with wind speed above 3 m s^{-1} could arise from: (i) an inadequacy of the physics in Fairall's formulation that is in common between day and night; (ii) an inadequacy in the value of the λ_0 and A parameters used here; (iii) biases at low wind speeds in the NWP fluxes used to drive the model (e.g., Horrocks et al., 2003a); (iv) a wind-related error in the skin SST retrieval; or (v) an error in the skin SST retrieval that depends on a factor correlated with wind speed (a confounding factor). A wind-related retrieval error would arise, for example, via an inadequate simulation of the dependencies of emissivity and reflected radiance (most likely in the forward view) when defining retrieval coefficients. Potential confounding factors are: inability of the retrieval to cope wholly with marine aerosol (whose concentration is correlated to wind speed); or systematic retrieval error related to atmospheric humidity, since column water vapour is correlated to wind speed globally.

Perhaps all of these factors play some role, but it seems likely that parameter error in the implementation of the Fairall model is significant. The implementation of the Fairall model used here was tuned in an earlier study at the Met Office for use with operational AATSR data, so as to minimize the wind-dependence of bias between operational AATSR and in situ buoys. Much of the residual wind-speed dependence seen here could therefore arise as an artefact of the difference between the operational and new ARC retrieval coefficients. A preliminary study suggests that the residual wind-speed dependence is reduced (data not shown) by removing the tuning step in the Met Office's implementation of the Fairall model and reverting to Fairall's original recommended formulation. This will preserve the independence of ARC SSTs from drifting buoy observations (see Embury and Merchant, 2012), and results will be reported in a later paper.

Table 4.4 shows the global retrieval biases and RSD for no skin correction, with skin correction

	Day	Day, $u > 3\,\mathrm{m\,s^{-1}}$	Night
$N_{matches}$	120916	104843	96857
None	-0.084 (0.235)	-0.084 (0.223)	-0.149 (0.227)
Donlon	0.122 (0.240)	0.112 (0.226)	0.061 (0.221)
Fairall	0.070 (0.241)	0.066 (0.229)	0.057 (0.228)

Table 4.4: Statistics of AATSR D2 sub-skin SST estimation for different skin effect corrections. Median of satellite-in situ, with RSD in parentheses, both in kelvin.

using the Donlon parameterization, and using the Fairall model. The Donlon formulation is effective at reducing both the night time bias and the RSD. While the Fairall correction does result in a slight positive bias (~0.05 K) it has brought the day and night retrievals into very close agreement. Use of the Fairall method also causes a small increase in RSD; equivalent to an independent noise signal of magnitude 0.04 K (night) or 0.06 K (day).

4.3.2 Thermal stratification

During the day solar heating of the upper ocean can cause the formation of a near-surface "warm layer" under low wind conditions (e.g., Lukas, 1991). As a result, the sub-skin temperature is warmer than temperatures at lower depths. In cases of sustained wind speed less than ~1 $\mathrm{m\,s^{-1}}$ and high incident solar radiation, the difference can exceed 5 K by early afternoon relative to the SST before dawn, as thermal stratification occurs. Such extreme events are seen in spring and summer in mid and high latitudes (Merchant et al., 2008a; Gentemann et al., 2008). More typically, the diurnal cycle of SST is of order a few tenths of kelvin.

To model diurnal warming for the ARC project we use the Kantha and Clayson (1994) model implemented at the Met Office, of which full details are given in Horrocks et al. (2003b). This model is a one dimensional turbulence closure model, giving a profile of modelled stratification throughout the model run. More highly parameterized models for diurnal stratification exist (e.g., Gentemann et al., 2009; Zeng and Beljaars, 2005), and these merit evaluation in future work.

Temperature differences are calculated between the sub-skin and depth SST at 0.2, 1.0, and 1.5 metres below surface. In 83% of cases these differences are <0.05 K. For conditions conducive to large diurnal warming events, the modelled stratification across the top ~20 cm can reach 0.5 K by 1000 h local time. Figure 4.5 shows the difference between the SST at 0.2 m estimated by adjusting AATSR D2 SST_{skin} for both skin effect and diurnal stratification, minus drifting buoy SST. Compared to Fig. 4.4, there is reduced difference between the day and night-time dependence of difference at low wind speeds, indicating that the stratification estimate is beneficial.

Table 4.5 shows how correcting for the diurnal stratification affects the global biases. While the thermal stratification correction has a visible improvement in the wind speed dependent biases shown in Fig. 4.5, it has negligible effect on the global biases and RSD. This is primarily due to the low incidence (14% below 3 $\mathrm{m\,s^{-1}}$) of extreme low wind speed conditions in the global dataset. Therefore, in Table 4.5, we also present statistics for cases of low thermal stratification (<0.05 K) and high thermal

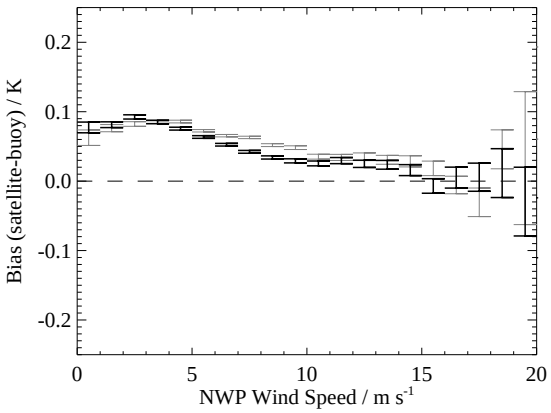

Figure 4.5: AATSR estimated $SST_{0.2\,m}$-drifters as a function of wind speed. Error bars span the range of median difference plus and minus one standard error.

	All cases	$\Delta T < 0.05$ K	$\Delta T > 0.05$ K
$N_{matches}$	120916	116763	4153
Sub-skin	0.070 (0.241)	0.069 (0.236)	0.150 (0.422)
0.2 m	0.066 (0.241)	0.068 (0.236)	-0.004 (0.441)

Table 4.5: Median and (in parentheses) RSD of satellite SST minus drifter SST in kelvin, where the satellite SST is the D2 SST skin retrieval adjusted to sub-skin and depth (20 cm). ΔT is the estimated stratification between the sub-skin and depth.

stratification (>0.05 K) separately. Day time cases of low modelled stratification have lower RSD than the high stratification cases. Using a harsher threshold of 0.01 K reduces the day time RSD to match the night-time RSD of ~0.23 K (these "extremely" low stratification cases cover 93% of all daytime matches). This is consistent, since stratification at night is usually expected to be negligible. The day-time cases with stratification greater than 0.05 K have warmer sub-skin SST estimates relative to drifters (by ~0.08 K) than the low-stratification cases, suggesting that truly stratified cases are being captured by this criterion. However, the average stratification correction for the high-stratification cases is 0.15 K, which indicates that on average the stratification is over-estimated.

The RSD of the day-time SST depth minus drifter SST is marginally greater than the RSD of day-time sub-skin SST minus drifter SST, for the more stratified sample. So, it seems that the sub-skin to SST depth adjustment is also adding some noise to these day-time cases. It is challenging to model the magnitude of the thermal stratification as it depends on both the wind speed and solar radiation history since dawn. It may be that the NWP data used to drive the model do not always capture these histories sufficiently accurately. Nonetheless, the stratification model is shown to have some skill at predicting the presence of stratification events and in reducing the bias of SST depth compared to drifting buoy SSTs.

Table 4.6 shows equivalent statistics for the GTMB array. Here the stratification effect is greater as the ATLAS moorings measure at a greater depth, and they are found in tropical locations where stratification events happen more often because of year-round high insolation and relatively low

	All cases	$\Delta T < 0.05$ K	$\Delta T > 0.05$ K
$N_{matches}$	6837	5783	1054
Sub-skin	0.072 (0.224)	0.055 (0.203)	0.245 (0.365)
1.0 m	0.049 (0.220)	0.051 (0.201)	0.030 (0.393)

Table 4.6: As Table 4.5 but for ATLAS moorings in the GTMBA. Measurement depth is 1m.

average wind speeds. The model is now reducing median and RSD satellite-in situ differences when considering both all matches and the low stratification events. For the more stratified cases, the stratification adjustment results in a small increase in RSD, but the significant improvement in median difference (from ~0.25 K to 0.03 K) shows that the stratification model is remarkably effective on average.

4.3.3 Diurnal cycle and time-adjustment

Sub-daily variability, particularly the diurnal cycle of SSTs, also leads to differences between the satellite SST and in situ measured SSTs if the two are not perfectly coincident in time. The maximum time-window allowed in the MDB is three hours, so up to three hours of solar heating or night-time cooling can occur between the two observations of the SST. Minnett (1991) showed that spatial separation between satellite and in situ of ~10 km and time separations of ~2 hours can introduce RMS errors of ~0.2 K into satellite validation. In this section, we show that, in fact, the errors associated with the time have a significant systematic component that can be compensated for. This is important where the distribution of satellite-buoy time difference is asymmetrical or where the satellite observation time is near the peak or trough of the diurnal SST cycle, because then the time difference effect can introduce an apparent bias in the validation results.

The effect of the time separation on retrieval bias is shown in Fig. 4.6. There is a cooling trend of 0.015 K/hour affecting night-time matches and a warming trend of 0.058 K/hour during the day. The Envisat satellite has a fixed 10:00 h Local Equator Crossing Time (LECT), and for most matches we can consider the local satellite observation times to be roughly 10 a.m. and 10 p.m. and therefore interpret Fig. 4.6 as showing the trends corresponding to two segments of the diurnal cycle. There is an upper limit to the daytime satellite-buoy retrieval bias for time differences >1.5 hours, which corresponds to local buoy times of around 8:30 a.m. and earlier, i.e. local times closer to the minimum of the diurnal cycle, before the sea surface begins to stratify significantly due to solar radiation.

The Fairall skin and Kantha-Clayson stratification models, discussed in the previous sections, can also be used to estimate the surface heating/cooling rates (that is, rates of change of surface temperature). This was originally performed in order to homogenise the ARC SST record – the Envisat satellite which carries the AATSR instrument has a 10:00 h LECT while the ERS-1 and ERS-2 satellites carrying the ATSR1 and ATSR-2 instrument have 10:30 h LECTs. This half-hour difference is sufficient to introduce a discontinuity in the day-time SST record of around 0.03 K (half this at night). Therefore, in order to develop a homogeneous record across the three sensors, an adjustment for the diurnal

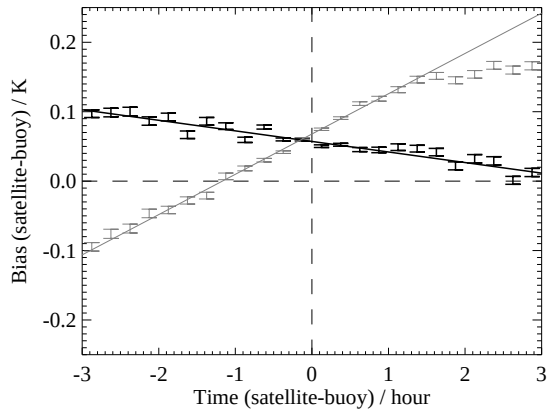

Figure 4.6: D2 $SST_{0.2\,m}$ retrieval bias as a function of satellite-buoy time difference for daytime (grey) and night-time (black) matches. Solid lines show linear best fit to data (only using time differences < 1.5 h for day time matches)

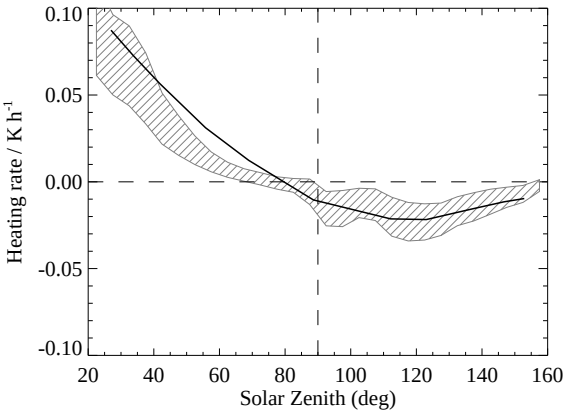

Figure 4.7: Rate of change of SST calculated from AATSR D2 $SST_{0.2\,m}$-drifter SST differences as a function of solar zenith angle. Line: average rate found for different solar zenith angle ranges calculated across the MDB. Shaded band: median ± RSD of rate simulated by the skin effect and stratification models for the same matches

cycle is required. Here we validate the modelled heating and cooling rates against observation-driven estimates.

Figure 4.7 shows the mean observed (line) and modelled (band) heating rates at the AATSR observation time (mostly close to 10 a.m. or 10 p.m.) as a function of solar zenith angle (SZA). The observed heating rates are calculated by linear regression of the satellite-buoy retrieval bias against time differences for a series of strata of SZA. The simulated heating rates are represented by the shaded band which gives the median ± RSD of simulated heating rate as a function of SZA for the same set of matches. The night-time rate (SZA > 90°) is between −0.01 and $-0.02\,K\,h^{-1}$ (i.e., modest cooling). The day-time (SZA < 90°) warming is strongly dependent on SZA, and is up to $0.1\,K\,h^{-1}$, consistent with the ~0.5 K mean amplitude of the daily cycle in the tropics and summer mid-latitudes.

As ocean surface mixing is driven by surface winds, the rate of heating will be strongly dependent

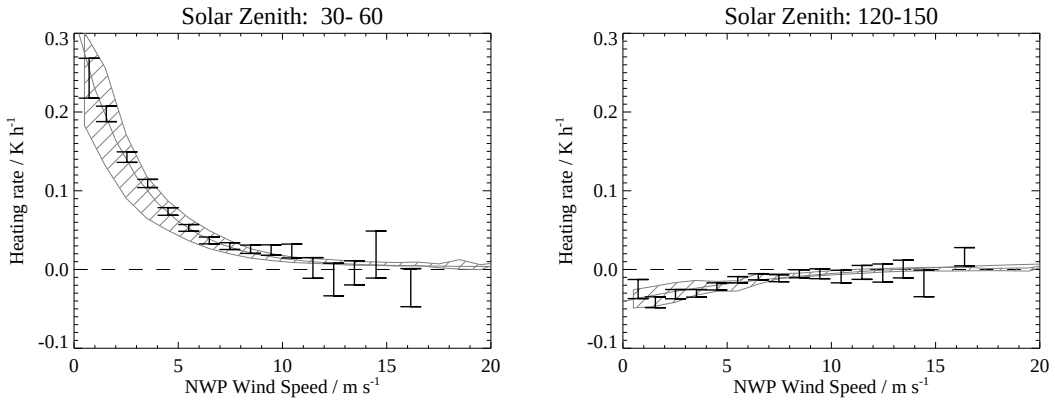

Figure 4.8: Surface heating rates as a function of wind speed calculated from observations (black bars) and simulation (shaded grey area). Left panel: subset of day-time matches with solar zenith angles between 30 and 60 degrees. Right panel, night-time matches with solar zenith angles between 120 and 150 degrees.

on wind speed in addition to solar zenith angle (insolation). This is shown in Fig. 4.8 where the heating rates are plotted against wind speed for two example ranges of SZA (one day and one night). There is very good agreement between the observed (black bars) and simulated (shaded area) heating rates. Within each range of SZA, the data show a dependence on wind-speed that is well described as an exponential function (solid line). So, an alternative to the model simulated time-corrections is an empirical model, comprising, for each solar zenith angle range, a fit against wind speed of form:

$$\frac{d\mathrm{SST}}{dt} = a_0 \exp(a_1 u) + a_2 \tag{4.2}$$

This empirical fit is less onerous to calculate than the simulated time corrections, and can also be applied to other observations with similar LECTs, including the ATSR1 and ATSR-2 MDBs (for which we have not modelled the half-hour SST time corrections). Figure 4.6 implies that the empirical model can be applied for the in situ observations up to 2 hours at least from the 10:30 h overpass time of ATSRs 1 and 2.

These temperature trends can affect both the apparent retrieval bias and noise in validation studies. For instance, if the distributions of satellite-buoy time differences were uniform and the trend applied exactly to all observations, the independent noise introduced would be:

$$\sigma = \frac{d\mathrm{SST}}{dt}\frac{\Delta t}{\sqrt{3}} \tag{4.3}$$

where Δt is the maximum time difference between satellite and buoy. For AATSR vs. drifter SSTs, the expected increase in SD/RSD is less than the above estimate because the time differences are not uniformly distributed. Larger time windows can also lead to bias artefacts if the satellite-buoy time differences are not symmetrically distributed around zero or span a range of local time over which the diurnal cycle is not linear, as is the case for the morning AATSR observations.

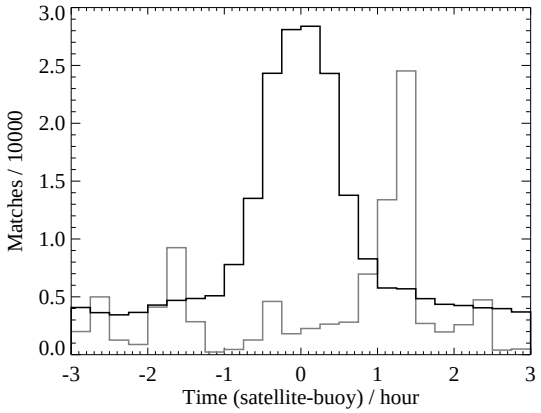

Figure 4.9: Distribution of satellite-buoy time differences for drifter (black) and GTMBA (grey) matches. Number of GTMBA matches has been scaled by 10.

	Day		Night	
Time Window	±3 hours	±1 hours	±3 hours	±1 hours
N. matches	120916	81829	96857	65963
None	0.066 (0.241)	0.069 (0.223)	0.057 (0.228)	0.056 (0.221)
Global mean rate	0.068 (0.234)	0.068 (0.221)	0.057 (0.228)	0.056 (0.221)
SZA-dependent rate	0.069 (0.232)	0.068 (0.220)	0.057 (0.228)	0.056 (0.220)
Wind and SZA dependent rate	0.068 (0.229)	0.068 (0.219)	0.057 (0.227)	0.057 (0.221)
Modelled	0.068 (0.230)	0.068 (0.220)	0.057 (0.228)	0.056 (0.221)

Table 4.7: Median and (in parentheses) RSD, in kelvin, of satellite $SST_{0.2\,m}$ minus drifter SST, for different methods of satellite-buoy time difference corrections indicated in leftmost column.

Figure 4.9 shows the distribution of satellite-buoy time differences in the AATSR drifter and GTMBA datasets. Most drifting buoys are making frequent reports (e.g., hourly) and as a result the majority, ~66%, of matches have time differences less than an hour. However, in the case of the GTMBA data only 23% of matches were within a one-hour time window, because of how frequently the ATLAS moorings transmit.

Table 4.7 shows the global statistics of $SST_{0.2\,m}$-drifter difference for AATSR D2 using different corrections for satellite-buoy time difference: (i) making no correction; (ii) using a single mean heating rate for day and cooling rate for night; (iii) using the SZA-dependent rates shown in Fig. 4.7; (iv) using the fitted exponential-in-wind-speed model for different SZAs (Eq. 4.2); and (v) using the simulated rates from the Fairall and Kantha-Clayson models. As the AATSR-drifter time differences are already low (see Fig. 4.9) the effects are relatively modest. Using the full correction reduces the daytime RSD from 0.241 K to 0.230 K (equivalent to removing independent noise of magnitude 0.07 K), which then matches the night-time RSD. A greater improvement in statistics is seen by limiting the data to matches within a one hour time window – day-time RSD reduces to 0.223 K. For comparison, the daytime matches with absolute time differences greater than 1 hour have a RSD of 0.286 K without correction for time difference, and 0.258 K with correction.

Table 4.8 shows the equivalent statistics for the tropical moored buoys. In this case there are

	Day		Night	
Time Window	±3 hours	±1 hours	±3 hours	±1 hours
N. matches	6837	1449	3081	832
None	0.049 (0.220)	0.105 (0.244)	0.021 (0.181)	0.049 (0.210)
Global mean rate	0.032 (0.197)	0.088 (0.248)	0.030 (0.177)	0.053 (0.208)
SZA-dependent rate	0.029 (0.199)	0.085 (0.248)	0.028 (0.179)	0.052 (0.209)
Wind and SZA dependent rate	0.030 (0.194)	0.084 (0.245)	0.028 (0.180)	0.051 (0.208)
Modelled	0.029 (0.196)	0.086 (0.240)	0.032 (0.179)	0.052 (0.204)

Table 4.8: Median and (in parentheses) RSD, in kelvin, of satellite $SST_{1.0\,m}$ minus GTMBA SST, for different methods of satellite-buoy time difference corrections indicated in leftmost column.

far fewer matches and the time differences are much more spread out due to the intermittent data reported though the GTS. Restricting the comparison to the smaller time window causes a significant increase in both RSD and bias due to the reduced sample size. When considering all matches within a three hour window, the time-difference correction reduces day-time RSD from 0.220 K to 0.196 K (equivalent to removal of 0.10 K of independent noise). There is also a positive effect in making day and night observations more consistent on average.

All the tropical moored buoy comparisons show better agreement (lower SD and RSD) with satellite measurement than do the drifting buoy data. In addition, there are clearly fewer outliers affecting the moored buoy comparison as the SDs are only slightly higher than RSDs, while the drifter SDs were approximately double the corresponding RSD. This is despite the tropical regions tending to have lower precision satellite SSTs due to the higher water vapour loading. The lower incidence of outliers seems, therefore, to reflect the higher quality of the tropical moored buoys compared to the drifting buoys in the data sets we use.

There is a trade-off between minimizing the satellite-buoy separation, in both time and space, and maximizing the number of matches to increase coverage and reduce sampling error. By matching in situ data to the closest ATSR pixel, we apply a spatial separation criteria of nominally 1 km, although it is effectively ~5 km due to the 0.1° resolution of the buoy location reports available to us. Increasing this space-window does not significantly increase the number of matches as the only "extra" ones are outside the satellite swath. When selecting the maximum temporal separation allowed we should consider how the SST can vary between two observations, which is driven by the diurnal cycle (or, sub-daily variability more generally), and the distribution of time separations. For example, a linear temperature trend and a symmetric distribution of time differences will give an unbiased satellite-in situ difference with only modest increase in validation RSD/SD.

The RMS of the simulated diurnal cycle is 0.03 $K\,h^{-1}$ at night and 0.10 $K\,h^{-1}$ during the day, which matches the 0.2 K difference for a two hour time separation identified by Minnett (1991). Simulations of the diurnal heating rates are in good agreement with the observed satellite-drifter differences. In the case of the AATSR-GTMBA matches, where the distribution of time differences is asymmetric with less than 25% of matches within a one hour window, correcting for the diurnal heating effect significantly reduced the satellite-in situ RSD while bringing the day/night differences into closer

agreement.

4.4 Validation

4.4.1 Global AATSR

In the previous section we use the satellite-in situ differences to explore and quantify the geophysical differences between the types of measurement. We now validate SST-depth estimates derived by adjusting retrieved skin SSTs to SST depth. The adjustments comprise the Fairall model for the skin effects, the Kantha-Clayson model for diurnal stratification, and time-difference effects parameterized in terms of solar zenith angle and wind speed, as described in previous sections. Performance of the new ARC retrievals are assessed partly in comparison with operational coefficients. Identical adjustments to SST depth have been applied, so that the effects of the different skin retrieval schemes can be clearly seen. The operational retrievals have already been shown to have relatively good performance. O'Carroll et al. (2008) performed a three-way uncertainty analysis between operational AATSR, in situ, and AMSR-E SSTs, and found the night time AATSR retrievals to have a lower error (standard deviation ~0.16 K) than in situ buoy measurements (0.23 K, although genuine geophysical variability may also contribute to this value).

As well as basic comparisons, we investigate whether the residuals vary systematically with factors which should affect the satellite retrieval but not the in situ measurements (such as atmospheric water vapour or satellite viewing geometry). Additionally, comparing the residuals from the different channel combinations and both the new ARC and the operational retrieval algorithms is informative about the characteristics of satellite retrieval uncertainties.

Note — we are not using the operational SSTs present in the ATSR multi-mission archive version 2.0, but are implementing the operational retrieval algorithm within the ARC processing chain which ingests Level 1b brightness temperature files. Both sets of SSTs (ARC and "ATS*") are generated from an identical set of input brightness temperatures using identical cloud screening etc., the only difference being the SST retrieval algorithm.

Table 4.9 shows global statistics comparing satellite-retrieved $SST_{0.2\,m}$ against drifter SST for both the operational and ARC AATSR retrievals. In all cases the ARC retrieval coefficients result in both a smaller median AATSR-drifter difference and lower RSD. With the operational SST there is a large spread in median AATSR-drifter differences with N2 more than 0.5 K warmer than D2. In practice, the operational nadir-only retrievals are used only in the absence of the D2 (during day) and D3 (at night) retrievals. There is a ~0.2 K difference between the dual-view retrievals, with D3 ~0.1 K warmer than drifters and D2 ~0.1 K cooler. The ARC SSTs are more consistent between types of retrieval: nadir-only N2 and N3 are ~0.03 K warmer than in situ while D2 and D3 are ~0.06 K warmer.

When considering the differences in RSD between the operational and ARC SSTs, the greatest improvement, from 0.45 K to 0.33 K, is seen for the basic N2 SSTs, followed by 0.29 to 0.23 K for the D2.

	N2	N3	D2	D3
Day (120916)				
ATS*	0.464 (0.431)		-0.086 (0.290)	
ARC	0.020 (0.317)		0.068 (0.229)	
Night (96857)				
ATS*	0.448 (0.450)	0.033 (0.200)	-0.108 (0.291)	0.088 (0.223)
ARC	0.031 (0.327)	0.034 (0.192)	0.057 (0.227)	0.055 (0.207)

Table 4.9: Global AATSR estimated $SST_{0.2\,m}$-drifter in kelvin, using both operational (ATS*) and new ARC retrieval coefficients.

This is due to these retrievals, which use only the 11 and 12 µm channels, being the most sensitive to atmospheric water vapour and therefore benefiting most from the improved ARC algorithm. For the night-only N3 and D3 the decrease in RSD is much lower ~0.01 K. However, the improvement may somewhat be masked because these RSDs are around 0.2 K, and are probably dominated by the random error of the buoy measurements.

Next, consider the differences in RSD between different algorithms. Largest RSD is obtained for N2 SSTs, reflecting the fact that single-view observations at 11 and 12 µm only do not have sufficient information content to deal with the full range of variability of atmospheric absorption. D2 SSTs improve significantly over N2 SSTs in terms of RSD, because the additional view significantly adds to the information content of the observations. The use of the 3.7 µm channel in the N3 SST also adds significant information compared to N2 SST, and N3 inversion does not amplify noise as much as D2 or D3 inversion (the magnitude of N3 coefficients is generally smaller than for the dual-view algorithms). N3 SSTs thus turn out to have the smallest RSD. Note, however, that the D3 SSTs still are less noisy than D2 SSTs. Moreover, the dual-view algorithms generally maintain accuracy better under unusual atmospheric conditions, and only they are highly robust (insensitive) to episodes of elevated stratospheric aerosol.

Table 4.10 shows global statistics comparing satellite estimated $SST_{1.0\,m}$ against in situ SSTs from the Global Tropical Moored Buoy arrays. Again, ARC SSTs all have lower RSD and more consistent median differences than operational SSTs. There are two important differences compared to Table 4.9. Firstly, the measurement error associated with the ATLAS moorings is much lower than drifters, and this is reflected in the lower RSDs for N3, D2, and D3 SSTs in Table 4.10. The ARC N3 SSTs now have a RSD of 0.128 K while the operational N3 SSTs have a RSD of 0.160 K. Secondly, the behaviour of the ARC N2 SSTs appears inconsistent with the previous comparison. While the ARC N2 SST - GTMBA does have lower RSD than the operational N2, it has a higher RSD than in the drifter comparison. This reflects the more limited geographical coverage of the GTMBA arrays which cover only tropical regions. These present difficulties for the split-window N2 retrievals due to the high levels of atmospheric water vapour.

	N2	N3	D2	D3
Day (6837)				
ATS*	0.399 (0.389)		-0.155 (0.246)	
ARC	-0.064 (0.369)		0.030 (0.194)	
Night (3081)				
ATS*	0.412 (0.424)	0.055 (0.160)	-0.155 (0.242)	0.073 (0.155)
ARC	-0.043 (0.379)	0.018 (0.128)	0.028 (0.180)	0.036 (0.145)

Table 4.10: Global AATSR-estimated $SST_{1.0\,m}$-GTMBA SST in kelvin, using both operational (ATS*) and new ARC retrieval coefficients.

4.4.2 Regional AATSR

Figure 4.10 shows the AATSR estimated $SST_{0.2\,m}$ - drifter SST against four different variables: latitude, NWP total column water vapour (TCWV), NWP wind speed, and across-track pixel position. ARC retrievals have improved inter-algorithm consistency, achieved independently of in situ observations by design (Embury and Merchant, 2012). ARC N3, D2, and D3 estimates are mutually consistent to within 0.05 K in agreement with the global statistics in Table 4.9. Latitudinal, TCWV and wind speed dependencies are also reduced.

In the latitudinal plots the greatest change between operational (Fig. 4.10b) and ARC retrievals (Fig. 4.10a) is seen for N2 and D2 where the operational retrievals have large systematic variations ~0.5 K. With the updated ARC retrievals the majority of these systematic variations are eliminated, with the exception of the N2 latitude bias pattern between 30S and 30N whose latitudinal structure remains. This is also apparent in the RSDs: in Fig. 4.10b the N2-drifter RSD is high (0.3–0.5 K) at all latitudes, but in Fig. 4.10a it is reduced to ~0.2 K in mid-latitudes while remaining high in the tropics. These retrieval issues are associated with the high atmospheric water vapour in the tropics which are particularly problematic for single-view split-window SST retrievals (Merchant et al., 2009). The latitudinal variation of the operational D2 SSTs is a known problem with AATSR retrievals – commonly shown in D2-D3 difference plots (Merchant et al., 1999) – and is an intrinsic property of the form of the linear retrieval used operationally (Merchant et al., 2006b). A latitude dependent bias correction has been devised (Birks, 2006), which has been applied to the operational D2 SST in Fig. 4.11, although it does not resolve the underlying problem which is variability with atmospheric state.

There are also two anomalies associated with Arctic retrievals. Firstly, the day-time N2 and D2 retrievals are ~0.5 K cooler than drifters north of 80° N. A similar effect is present in the operational SSTs, but less obvious due to other zonal variations. However, this value is based on only 24 matches to two separate drifters. It is difficult to assess, therefore, whether the biases arise from retrieval problems related to extreme atmospheric conditions, prevalent cloud contamination, or other factors. Secondly, there is a ~0.1 K difference between night-time D2 & D3, and the other retrievals within the Arctic Circle. This may be due to a cloud screening failure – the nadir-only and day-time retrievals are all consistent, while the dual-view night-time retrievals are warmer in comparison. Such an effect would occur if there is a small amount of undetected cloud in the forward view.

Figure 4.10d shows the large systematic variability of the operational N2 and D2 retrievals as

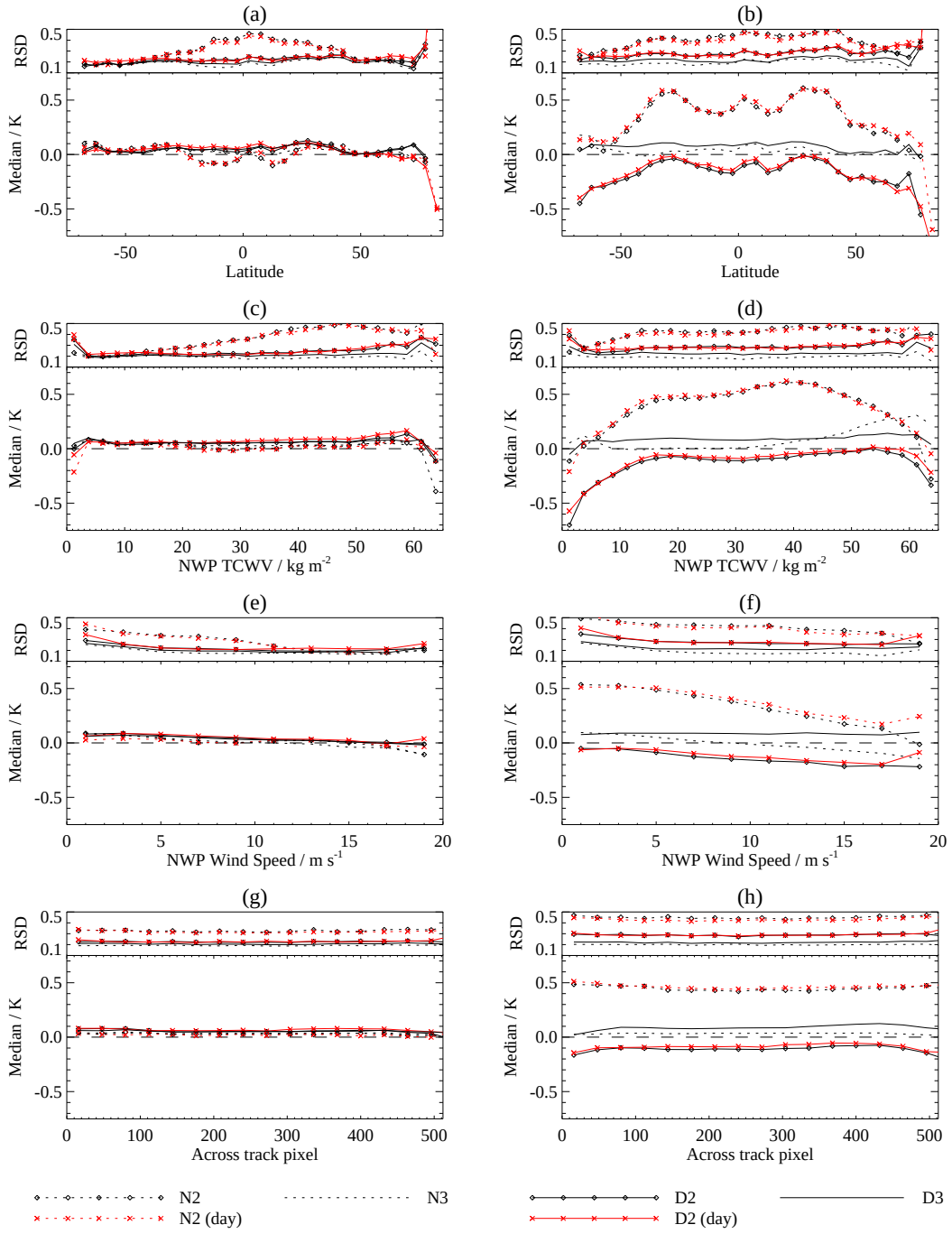

Figure 4.10: AATSR estimated $SST_{0.2\,m}$-drifters as a function of latitude (a,b), TCWV (c,d), wind speed (e,f), and pixel position (g,h). Left column (a,c,e,g) shows ARC retrievals, right column (b,d,f,h) shows operational retrievals. Dashed with symbol - N2; dashed - N3; solid with symbol - D2; solid - D3. Black indicates night-time data; red (with X symbol) indicates day-time data.

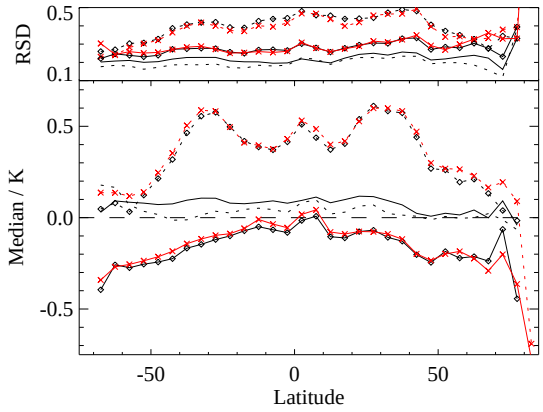

Figure 4.11: As Fig. 4.10b, but including the latitudinal-dependent offset to the D2 SST retrieval (see main text).

a function of TCWV. The N3 retrieval also shows a dependence on TCWV over 30 kg m^{-2} reaching ~0.3 K for the wettest atmospheres. Using the ARC retrievals (Fig. 4.10c) eliminates the majority of this dependence with the exception of TCWV values less than 5 kg m^{-2} or greater than 60 kg m^{-2}. Retrieval-drifter differences are 0.1–0.2 K colder for these extremes. As shown in Embury and Merchant (2012), this is partly expected given the retrieval design which leads to inherent biases at the extremes of the atmospheric conditions used to generate the coefficients. For example, the ARC D2 retrieval is expected from simulation to have a cold bias of ~0.2 K for TCWV ~65 kg m^{-2}; N3 and D3 are, in simulation, less affected and the N2 more so, as observed in practice. There may also be some bias because extremely humid atmospheres are expected to be associated with less reliable cloud masking with a higher chance of residual cloud affecting brightness temperatures used for SST retrievals and causing cold biases.

For moderate TCWV values, the ARC N3, D2, and D3 retrievals show a slight linear trend from ~0.05 K for dry conditions to ~0.1 K for wet conditions – comparable to the operational D3 retrieval. While the ARC N2 retrieval does match the others for dry conditions, it diverges slightly for intermediate values of TCWV (20–40 kg m^{-2}) with N2-drifter differences around 0.0 K. At higher values of TCWV the N2-drifter differences increase closer to the other retrievals. However, when viewing a sufficiently opaque atmosphere, the retrieval relies more on prior information (the climatological correlation of SST and lower atmospheric temperatures) with decreased sensitivity to the observations (Merchant et al., 2009). This can also be seen in the N2-drifter RSD, which increases with TCWV, and is little different from the operational N2-drifter RSD above about 40 kg m^{-2}.

Wind speed dependence is shown in Fig. 4.10e and f. The operational retrievals all have widely different sensitivities to the wind speed, with N2 the most sensitive and D3 showing no dependence. With the ARC coefficients all four retrievals have the same small dependence on wind speed seen in subsection 4.3.1. The consistency of the residual dependence between the four different ARC retrievals is encouraging. As discussed in subsection 4.3.1, it is likely that a significant part of the residual

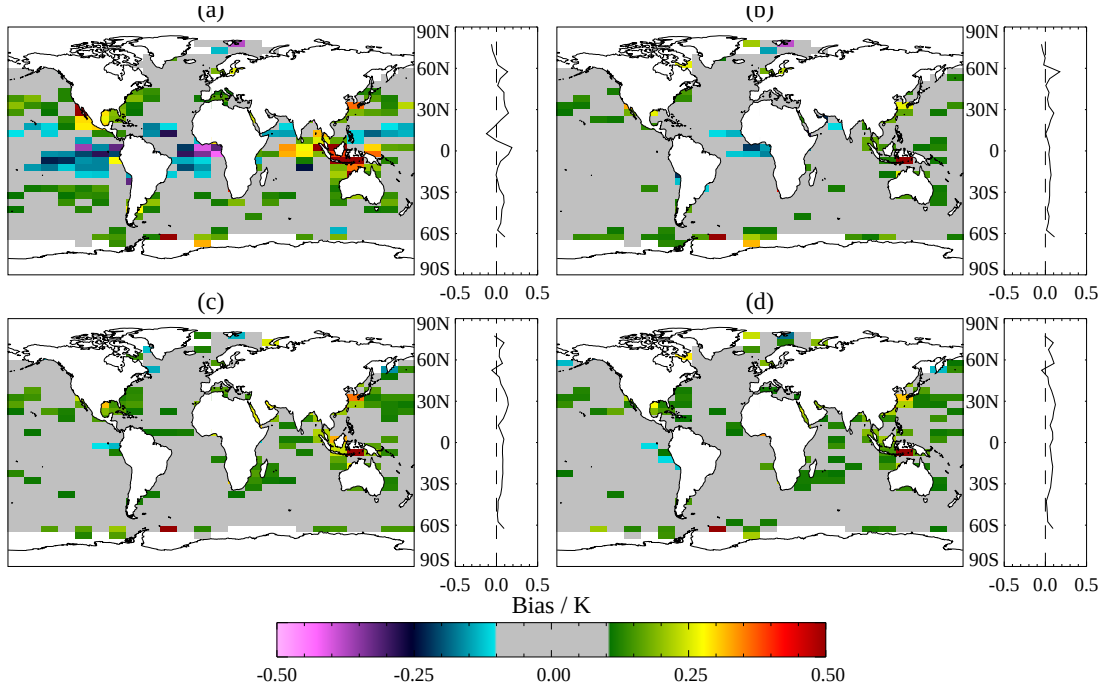

Figure 4.12: Median of difference between AATSR-estimated $SST_{0.2\,m}$ and drifting buoy SST for (a) N2, (b) N3, (c) D2, and (d) D3 retrievals using new coefficients.

dependence is an artifact of the skin-effect adjustment, implying that the wind-speed dependence of the new retrievals is yet smaller.

Figure 4.10g and 4.10h show the retrieval-drifter differences as a function of across track pixel position. Systematic biases in the operational retrievals across track arise from assuming the satellite viewing geometry is a fixed function of the across track pixel position. The N2 retrieval-drifter difference is approximately 0.1 K lower in the centre swath than at the edges, while the D2 and D3 retrievals show a slight "m" shape similar to the simulations shown in Embury and Merchant (2012). This viewing geometry dependence may also be contributing to the north-south asymmetry affecting the operational D3 (Fig. 4.10b). The ARC scheme removes these across track biases.

Figure 4.12 and 4.13 show global distributions of the median and RSD of difference between AATSR-estimated $SST_{0.2\,m}$ and drifter SST. The N3, D2, and D3 retrievals all have good agreement with the in situ data with cell-mean differences mostly <0.1 K and RSDs <0.2 K. The largest satellite-drifter differences are around Svalbard, Indonesia, and in the Southern Ocean and these are all locations where satellite SST retrievals are challenging and there is poor in situ coverage. In all three cases the number of in situ matches is very low (see Fig. 4.2) as is the number of individual buoys making the reports. The N2 retrieval, however, has a much more varied regional bias pattern, especially along the equator following the high equatorial RSD and zonal differences seen in Fig. 4.10a. Here we see the equatorial satellite-drifter differences are negative in the Pacific and Atlantic Oceans, but positive in the Indian Ocean and around Indonesia. Similarly the N2 RSD is high at ~0.5 K throughout most of the tropics.

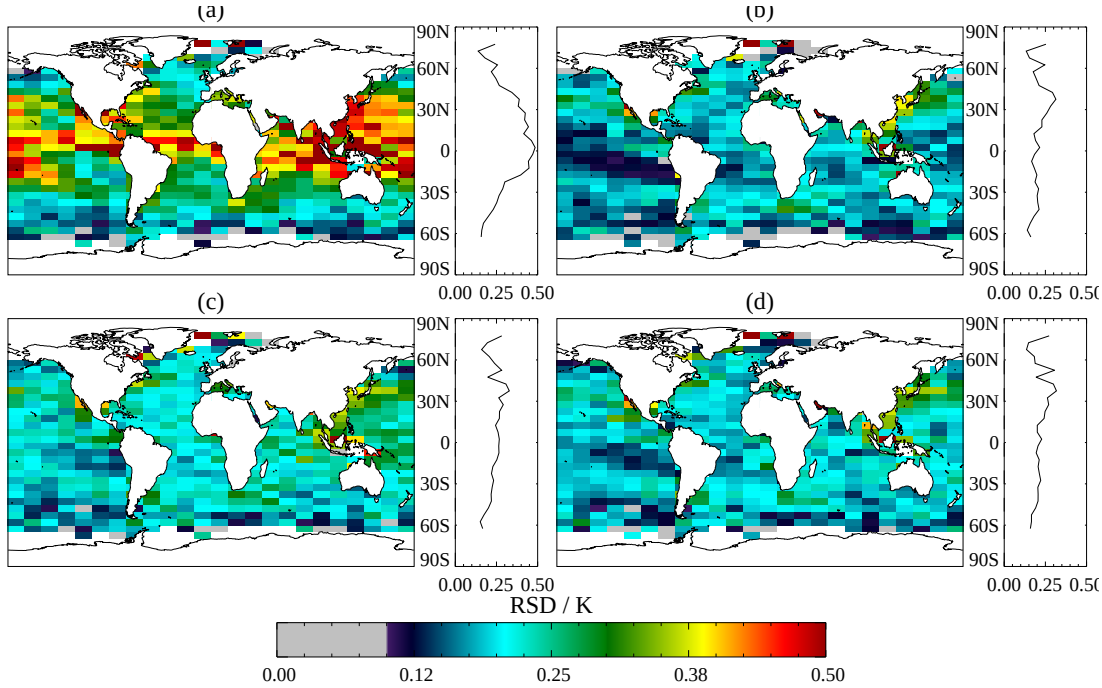

Figure 4.13: Robust standard deviation of difference between AATSR-estimated $SST_{0.2\,m}$ and drifting buoy SST for (a) N2, (b) N3, (c) D2, and (d) D3 retrievals using new coefficients.

The degree of statistical significance of the median difference of each cell from zero was evaluated using Student's t test, assuming that all matches have independent errors. (Correlated errors between different matches to a given buoy mean that the significance levels found in this way are over-estimates.) The test indicated that the calculated median differences are statistically significant at the 90% confidence level for 65 to 70% of latitude-longitude cells depending on the type of retrieval. Over half (50–55%) are significant at the 99% level. In this context, statistically significance indicates that the median satellite-drifter difference is non-zero, but these differences are still mostly less than 0.1 K except for the N2 retrieval. Cells where the observed median differences are not statistically significant include the three areas of low in situ coverage and large (~0.5 K) satellite-drifter differences discussed previously. The remaining cells which are 'not significant' correspond to cases where the regional biases are very low (~0.01 K for the N3, D2 and D3 retrievals).

One of the targets for ARC (Merchant et al., 2008c) was for regional biases to be within 0.1 K. On the basis of Fig. 4.12, this objective has been met for AATSR dual-view SSTs relative to drifting buoys across 80% of the global ocean.

4.4.3 ATSR-2

Table 4.11 shows global statistics comparing satellite estimated $SST_{0.2\,m}$ against drifter SST for ATSR-2. As with the AATSR results (c.f. Table 4.9), the ARC retrieval coefficients result in a lower satellite-drifter RSD and more consistent median differences. However, the satellite estimates of SST are now ~0.14 K warmer than drifters or ~0.1 K warmer than the corresponding AATSR comparison. Furthermore

	N2	N3	D2	D3
Day (46017)				
ATS*	0.574 (0.438)		-0.004 (0.357)	
ARC	0.064 (0.350)		0.138 (0.315)	
Night (36461)				
ATS*	0.575 (0.446)	0.129 (0.238)	-0.008 (0.335)	0.133 (0.260)
ARC	0.095 (0.351)	0.137 (0.235)	0.143 (0.291)	0.143 (0.250)

Table 4.11: Global ATSR-2 estimated $SST_{0.2m}$-drifter in kelvin, using both operational (ATS*) and new ARC retrieval coefficients.

	N2	N3	D2	D3
Day (6097)				
ATS*	0.525 (0.381)		-0.103 (0.263)	
ARC	0.045 (0.358)		0.121 (0.243)	
Night (1675)				
ATS*	0.586 (0.438)	0.208 (0.176)	-0.071 (0.265)	0.123 (0.160)
ARC	0.134 (0.407)	0.181 (0.134)	0.158 (0.238)	0.158 (0.156)

Table 4.12: Global ATSR-2 estimated $SST_{1.0m}$-GTMBA in kelvin, using both operational (ATS*) and new ARC retrieval coefficients.

the N2 is ~0.05 K cooler than the other retrievals, a much larger discrepancy than seen for AATSR. Table 4.12, comparing satellite estimated $SST_{1.0m}$ against in situ SSTs from the Global Tropical Moored Buoy arrays, again shows while the ARC SSTs are more consistent in satellite-buoy differences, the ARC SSTs are significantly warmer ~0.16 K than in situ data. The RSDs for the ATSR-2 N3 and D3 estimates shown in Table 4.12 are only slightly higher ~0.01 K (equivalent to ~0.05 K independent noise) than the corresponding AATSR values in Table 4.10, while for the D2 RSD the ATSR-2 value is ~0.06 K higher (equivalent to 0.16 K independent noise) than AATSR.

Figure 4.14 is the ATSR-2 equivalent to Fig. 4.10 and shows the ATSR-2 estimated $SST_{0.2m}$ - drifter SST. The same comments made concerning Fig. 4.10 apply here too, though the satellite-drifter differences are ~0.1 K warmer for the ARC ATSR-2 estimates than the AATSR ones. The ATSR-2 SST estimates are more sensitive to TCWV than the AATSR estimates, in Fig. 4.14c, the satellite-drifter differences go from ~0.1 K for dry atmospheres to ~0.2 K for wet (values were ~0.05 K and ~0.10 K for AATSR). This also shows in the latitudinal differences, Fig. 4.14a, where the ATSR-2 tropical estimates are ~0.2 K warmer than drifters. Behaviour of RSD with latitude and TCWV is also different for ATSR-2. While N3 and D3 RSDs were almost independent of TCWV with AATSR, they are now higher (~0.05 K) for low TCWV and decrease to close to the AATSR values for high TCWV; D2 RSDs, however, are elevated regardless of the TCWV value.

4.5 Discussion

In the ARC project we aim to retrieve skin SSTs from the satellite observations, and also to generate SST estimates at depths representative of buoy observation depths at a standardized local time. Therefore, in this initial validation, we have considered in detail the relationships between skin SST retrieved

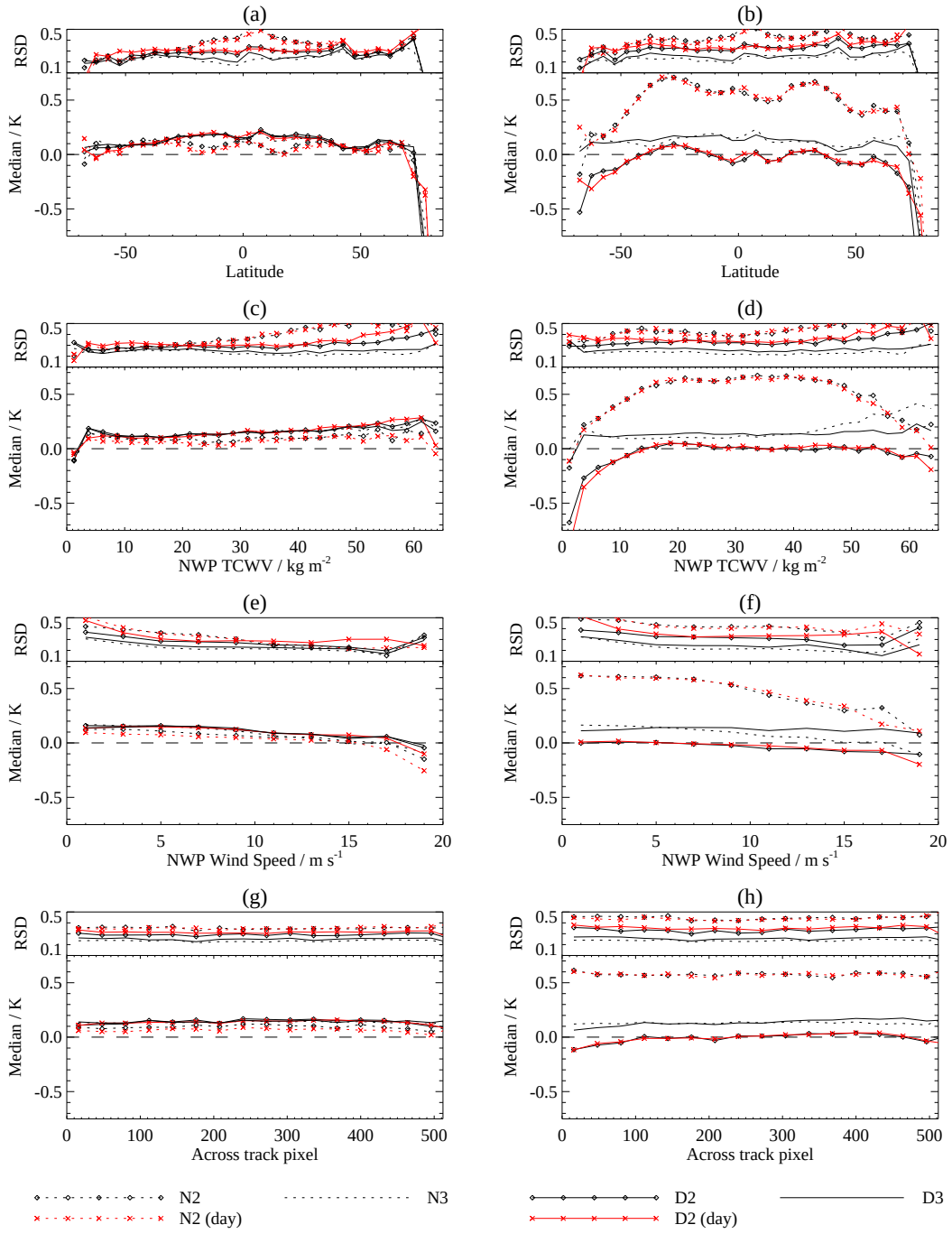

Figure 4.14: ATSR-2 estimated $SST_{0.2m}$-drifters as a function of latitude (a,b), TCWV (c,d), wind speed (e,f), and pixel position (g,h). Left column (a,c,e,g) shows ARC retrievals, right column (b,d,f,h) shows operational retrievals. Dashed with symbol - N2; dashed - N3; solid with symbol - D2; solid - D3. Black indicates night-time data; red (with X symbol) indicates day-time data.

from AATSR and in situ observations, with respect to the skin effect, near-surface stratification and the diurnal heating and cooling cycle.

To estimate the skin effect we use the Fairall et al. (1996) model tuned for use with the operational AATSR D3 retrievals in a previous study. Without accounting for the skin effect, there is apparently a ~0.07 K discrepancy between day and night-time SST retrievals relative to in situ observations. Our interpretation is that this relative bias is the average mid-morning modification of the skin effect by insolation. Correction with the Fairall skin model (which accounts for fluxes into the skin layer) reduces the day-night discrepancy to ~0.01 K.

The dependence of the Fairall skin effect model with wind speed is not entirely consistent with the data, with a wind-speed dependent residual of 0.1 K at about $3\,\mathrm{m\,s^{-1}}$ falling close to zero by $10\,\mathrm{m\,s^{-1}}$. While having a positive effect on bias, applying the Fairall model does slightly increase the scatter of the resulting $\mathrm{SST_{subskin}}$-drifter differences. There are various possible reasons for these inconsistencies. There may be wind-speed related biases in the skin SST estimated from AATSR, either directly via inadequately modelled surface emissivity or via confounding of wind-speed with some other factor that causes bias. Moreover, the greater noise apparent in the Fairall model estimates may arise from error in the input fluxes rather than limitations of the model. However, we think the most likely reason is that the Fairall model parameters used in this study were originally intended to use with operational AATSR D3 retrievals. The residual wind-speed dependence seen here is likely an artefact of the different wind-speed sensitivity of the operational $\mathrm{SST_{skin}}$ retrievals.

There are two factors which suggest the ARC D3 skin SSTs are less likely to be in error than the operational D3 skin SSTs. Firstly, in validation all the operational retrievals have widely different responses to wind speed while the ARC retrievals are all consistent. Secondly, the independent night-time skin correction of Donlon et al. (2002) has a wind-speed dependence that is consistent with the ARC observations, and does reduce the scatter of $\mathrm{SST_{subskin}}$-drifter differences when applied to night-time data.

The second step involved in adjusting $\mathrm{SST_{subskin}}$ to drifter-depth is to estimate near-surface stratification. This is based on running a Kantha-Clayson model forced by the time-history of NWP fluxes from the previous dawn (or time of minimum solar elevation). Diurnal stratification within the uppermost few tens of cm of the ocean around 10 a.m. local time is generally modest (<0.1 K on average) but is detectable in the observations for wind speeds less than $\sim 3\,\mathrm{m\,s^{-1}}$. The Kantha-Clayson model we use for thermal stratification does reconcile daytime and nighttime satellite-buoy differences for low wind speeds ($<3\,\mathrm{m\,s^{-1}}$) to ~0.02 K. Alternatively, users of ARC SSTs for climate applications may prefer to exclude potential thermal stratification events based on the model predictions: we find that this approach also reduces the day-night differences and reduces the daytime RSD from 0.24 K to match the night-time value of 0.23 K.

The final step in adjusting skin to drifter-depth SST accounts for the diurnal cycle that is associated with warming during the morning overpass and cooling during the evening overpass. There are two reasons to address this. First, the uncorrected diurnal cycle increases scatter in validation against

observations collected over a time window around the satellite time. Moreover, if satellite-buoy time differences are asymmetrically distributed or include the time of minimum or maximum diurnal cycle, the effect can introduce bias artefacts between day and night satellite SSTs. While these problems can be avoided by using very harsh matching criteria between satellite and in situ (i.e. time differences <1 hour or better) this can drastically reduce the number of matches available for validation. Second, there is a difference in overpass time of about 30 minutes between AATSR and ATSRs 1 and 2, which must be adjusted for in an SST time series for climate applications.

We have explored heating and cooling rates around the satellite observation time both in simulation (using the Kantha-Clayson model referred to above) and empirically. We determined the local surface heating (or cooling) rate during the AATSR overpass from observed time-dependence of satellite-in situ differences, as a function of solar zenith angle and surface wind speed. The observed heating rates show excellent functional agreement with those calculated with the Kantha-Clayson model.

Because we have a large validation dataset in which most buoys reports are within an hour of the satellite observation time, correcting for the satellite-buoy observation time difference in the AATSR-drifter match-up data base results in only slightly improved RSDs. However, the AATSR-GTMBA matches are limited in number and not so favourably sampled in time, and correcting for the time-difference results in significantly reduced RSDs and better day/night bias agreement, while using harsher matching criteria results in higher RSDs and worse biases.

Our initial validation results for ARC $SST_{0.2\,m}$ from AATSR against drifter measurements is encouraging. Global biases are around 0.03 K for nadir-only retrievals and 0.05 K for dual-view retrievals. Scatter is also low with RSD values of 0.33, 0.19, 0.23, 0.21 for the N2, N3, D2, and D3 algorithms respectively. However, there is a limitation to the validity of these RSD results. As shown by O'Carroll et al. (2008), quality controlled drifting buoys reported on the GTS themselves have an error of ~0.2 K, so the RSDs of the ARC-drifter comparisons may predominantly reflect the quality of the in situ data rather than the SST retrievals. N3 and D3 SSTs in particular are expected (from simulation) to have scatter markedly less than 0.2 K (Embury and Merchant, 2012). Using the GTMBA dataset based on the more accurate ATLAS moorings, RSDs of 0.38 (N2), 0.13 (N3), 0.18 (D2), and 0.15 (D3) are obtained, which are a better estimation of the point scatter error attributable to the ARC SST depth estimates. The SST depth estimates are formed by adding adjustments for skin and stratification effects to the primary ARC observations, which are skin SST retrievals. With some limitations that have been noted, these adjustments seem effective for systematic aspects of skin-depth SST difference. However, the adjustments tend to add a small amount of scatter, since they have random uncertainty of their own (e.g., see Table 4.4). The RSDs quoted above therefore give us upper limits to the random uncertainty inherent in the ARC SST_{skin} estimates.

Systematic biases have been observed for previous (operational) ATSR-series SSTs against latitude, TCWV, wind speed and viewing geometry. The magnitude of these systematic effects has been reduced using ARC SSTs to < 0.1 K for all except the most extreme regimes, namely, where TCWV <5 $kg\,m^{-2}$,

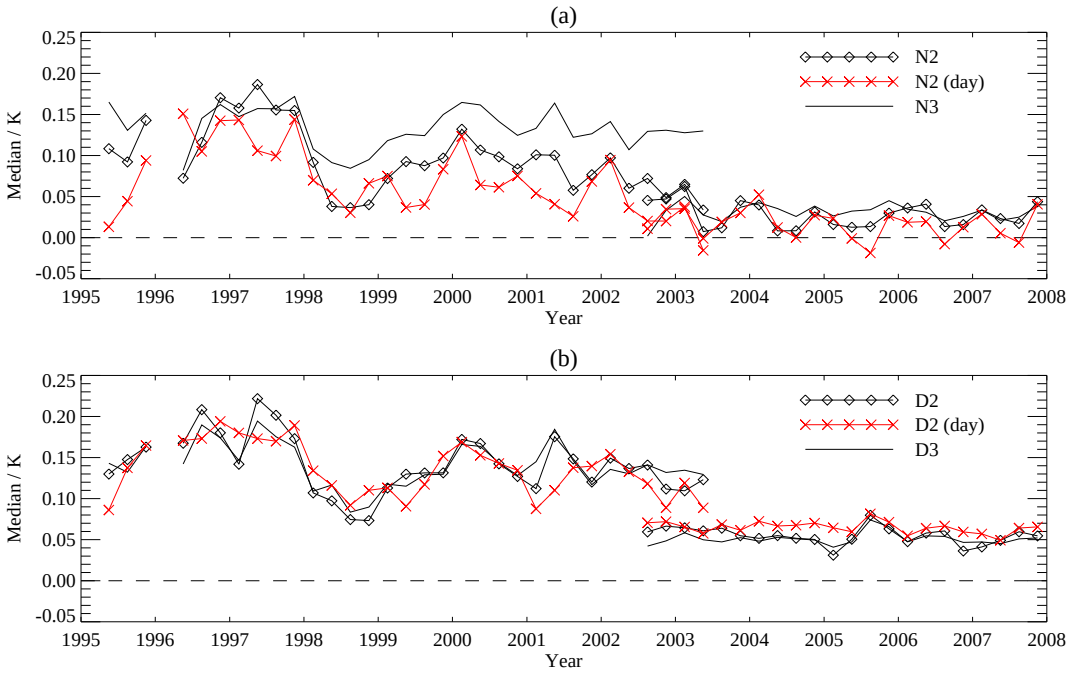

Figure 4.15: Time series of ATSR-2 (up to mid-2003) and AATSR (from mid-2002) estimated $SST_{0.2\,m}$-drifters. (a) Nadir-only retrievals: solid with symbol N2; solid N3. (b) Dual-view retrievals: solid with symbol D2; solid D3. Red lines (with X symbol) indicates day-time data.

TCWV >60 kg m^{-2}, and for latitudes polewards of about 65°. Regional validation show $SST_{0.2\,m}$-drifter differences are <0.1 K for the N3, D2, and D3 retrievals for 80% of the area of the global ocean. The exceptions are high-latitudes and Indonesian waters, both of which have very sparse in situ coverage. The N2 is the only retrieval with significant regional SSTs biases exceeding 0.1 K; the greatest variations are seen in the tropics, where N2 SSTs are several tenths of a degree cool in the Pacific and Atlantic, and several tenths of a degree warm in the Indian Ocean and around Indonesia. These patterns have some resemblance to those described for AVHRR in Merchant et al. (2009) and it should be noted that the limitations of the Pathfinder "split-window" SSTs discussed in that paper will apply to some degree also to ARC N2 SSTs, despite the ARC coefficients being banded by TCWV.

Initial validation of ARC ATSR-2 SSTs show a reduction in functional and regional biases comparable to that found for AATSR. Likewise, there are also reduced RSDs compared to operational algorithms. The global $SST_{0.2\,m}$-drifter differences, while more consistent than the operational retrievals, are around 0.1 K warmer than their AATSR equivalents resulting in a median $SST_{0.2\,m}$-drifter difference of ~0.14 K. However, as shown in Fig. 4.15 the ATSR-2 $SST_{0.2\,m}$-drifter differences change significantly over time. All the ATSR-2 retrievals are somewhat warmer, ~0.15 K, than drifters during the early part of the ATSR-2 mission (1996–1998) but the N2 retrieval in particular cools towards the end of the ATSR-2 mission. During the overlap between the two satellites only the N3 and D3 retrievals show the ~0.1 K difference between ATSR-2 and AATSR, for D2 the difference is reduced to ~0.05 K, and N2 is <0.01 K.

There are several causes which could be contributing to the ATSR2-AATSR discrepancy. It is most likely to arise from calibration difference between the two instruments. Given that the process of SST retrieval tends to amplify any calibration errors (Merchant and Le Borgne, 2004), this is quite plausible, and could arise by chance even if all channel calibrations were accurate to <0.1 K. Furthermore a change in calibration over time could lead to the inter-algorithm differences seen with ATSR-2. Another cause (which we consider less plausible) is inhomogeneity in the ARC processing inputs over time. The ARC processing uses prior NWP information for both cloud detection and to determine the TCWV that selects the SST coefficients used. The NWP fields are taken from ERA-40 and, after the end of the ERA-40 timeseries in August 2002, ECMWF-operational analyses. We consider this potential cause less plausible for a 0.1 K offset, since neither cloud detection nor SST retrieval are strongly responsive to the NWP state. Lastly, it could be an artefact of changes in the in situ record it self. The number of drifter matches in 2000 was double that in 1997. This can affect ATSR2-drifter differences if the new additions are altering the geographical distribution of the buoys as the ATSR-2 SST retrievals have a greater regional variability (~0.1 K) than the AATSR ones. As part of future work to ensure homogeneity of the ARC SST record a detailed overlap analysis will be performed using common in situ data so that ATSR-2/AATSR SSTs (and brightness temperatures) can be directly compared, and ATSR-2/AATSR SSTs will be compared against quality controlled in situ data selected for high stability.

This initial validation has established a high degree of consistency between ATSR skin SST retrievals, day and night, for ATSR-2 and AATSR, over most of the regimes of the global oceans. The subtle variability of the ocean thermal skin and diurnal stratification has been estimated to adjust the ATSR skin SST retrievals to form SST_{depth} estimates, which are more directly comparable to the buoy observations available for validation. We have been able to validate the systematic behaviour of the skin and stratification models used, and have shown that ATSR-2 and AATSR SST_{depth} can be adjusted to a common satellite overpass time with good confidence, thereby minimizing aliasing of the diurnal cycle into the multi-mission record of SST. This all represents significant progress towards an independent "climate quality" (Merchant et al., 2008c) SST record capable of refining our knowledge of marine climate change over recent decades.

4.6 Acknowledgements

The ATSR Reprocessing for Climate project is jointly funded by the Natural Environment Research Council, the Ministry of Defence, and the Department of Energy and Climate Change (previously Defra). Archived ATSR data and computer facilities are provided by the NERC Earth Observation Data Centre (http://www.neodc.rl.ac.uk) and we would like to thank Matt Pritchard for his work on the multi-mission archive and his assistance with the ARC project. We would also like to thank John Kennedy for access and support for the Met Office in situ SST dataset, and Karsten Fennig for adapting the Met Office SST skin-to-depth correction code for use in the ARC project.

Chapter 5

Harmonisation of Satellite Datasets into a single record

Owen Embury, Christopher J. Merchant

Abstract

Clear-sky brightness temperatures (BTs) observed by the series of three Along Track Scanning Radiometers show differences between the three sensors. Using radiative transfer simulation, we show these differences are only partly explained by reported sensor characteristics (spectral response function differences). The residual differences are inferred to arise from sensor calibration and/or radiative transfer simulation errors. For the most part, the residual differences are <0.1 K, but, uncorrected, these lead to discontinuities in retrieved sea surface temperature (SST) estimates of order ~0.1 K between sensors. Harmonisation of the record is required to create a stable climate data record, and we give reasons to prefer to harmonise BTs rather than SSTs. Differences of simulated minus observed BT (double differences) between each pair of instruments are obtained during overlaps between missions. The BT adjustments needed to harmonise the record with respect to radiative transfer simulation are inferred. Updated SST retrieval coefficients are then generated which produce the harmonised SST record. Comparing the harmonised SSTs against tropical moorings shows that the two independent datasets are stable relative to each other from 1995 to 2012 to within ±2 mK year^{-1} at 95% confidence.

5.1 Introduction

Accurately determining the rate at which the global sea surface temperature (SST) is changing is of key importance for climate change detection studies. In order to achieve this it is necessary to have a SST record that is both stable, i.e. constant in the absence of real SST changes, and sufficiently long to

detect climatic trends in the presence of inter-annual variability. Allen et al. (1994) estimate that a minimum of 10–15 years is required for trend detection. When SST records are based on satellite data it is necessary to merge data from multiple sensors in order to achieve this length of record.

The Along Track Scanning Radiometer (ATSR) Reprocessing for Climate (ARC) project produced a global SST dataset spanning 1991 to 2009 based on satellite observations intended primarily for climate research (Merchant et al., 2012). This has subsequently been extended within the European Space Agency Climate Change Initiative (CCI) to the end of life of the Advanced ATSR (AATSR) in 2012. Merchant et al. (2008c) discuss the objectives of the ARC project which include: a minimum length of 15 years, independence of in situ data, stability better than 5 mK year^{-1}, and high accuracy (~0.1 K). These objectives match well the statements of requirement for climate observation set out by the Global Climate Observing System (GCOS, 2011). Previous papers have described the SST retrieval algorithm (Embury and Merchant, 2012) and its basis in radiative transfer (RT) simulations (Embury et al., 2012a), and a preliminary comparison against in situ data (Embury et al., 2012b). This paper presents the method used to harmonise the SST dataset across the three ATSR instruments contributing to the record and achieve the stability target.

Embury et al. (2012b) showed that the unharmonised SSTs produced within the ARC project were of good quality with low global biases (<0.05 K for AATSR and ~0.14 K for ATSR-2) and low variability of regional biases (<0.1 K). However, an inconsistency between sensors of ~0.1 K is unacceptably large in the context of stability requirements for an SST climate data record.

The degree of inconsistency is illustrated in the time series of unharmonised $SST_{0.2\,m}$ minus in situ drifter measurements shown in Fig. 5.1 — an updated version of Figure 15 from Embury et al. (2012b) with the addition of ATSR-1 data and an additional 4½ years of AATSR data. While the AATSR data from mid-2002 onwards appears stable with close agreement between the different channel combinations, the earlier data are more divergent. In the ATSR-2 data between 1995 and 2003, the nadir-view three-channel (N3) combination, which includes the 3.7 µm channel, is up to ~0.1 K warmer than the nadir-view two-channel combination which does not use the 3.7 µm channel. The difference between the different retrievals changes over time by about 0.05 K. This time dependence may not be due (or not solely due) to drift in instrument calibration: bias varies geographically by ~0.1 K and there was significant evolution of the drifter network during this period, from limited coverage in the early 1990s to near-global coverage in the early 2000s.

Further divergence can be seen in the ATSR-1 data between 1991 and mid-1996. Firstly, the nadir-only retrievals (Fig. 5.1a) are biased by the presence of stratospheric sulphate aerosol from the eruption of Mount Pinatubo in 1991 with large negative biases up to 0.8 K seen in 1992 and decaying through 1993. This is expected: the nadir-only retrievals are not aerosol robust (Merchant et al., 1999) and not recommended for use within climate applications during this period. (For a discussion about which channel combinations are most suitable for different applications, see Merchant et al., 2012). There is a negative excursion of 0.1 K in the dual-view retrievals (Fig. 5.1b) between the start of the record and the end of 1992. This is thought to indicate a residual sensitivity to stratospheric aerosol in

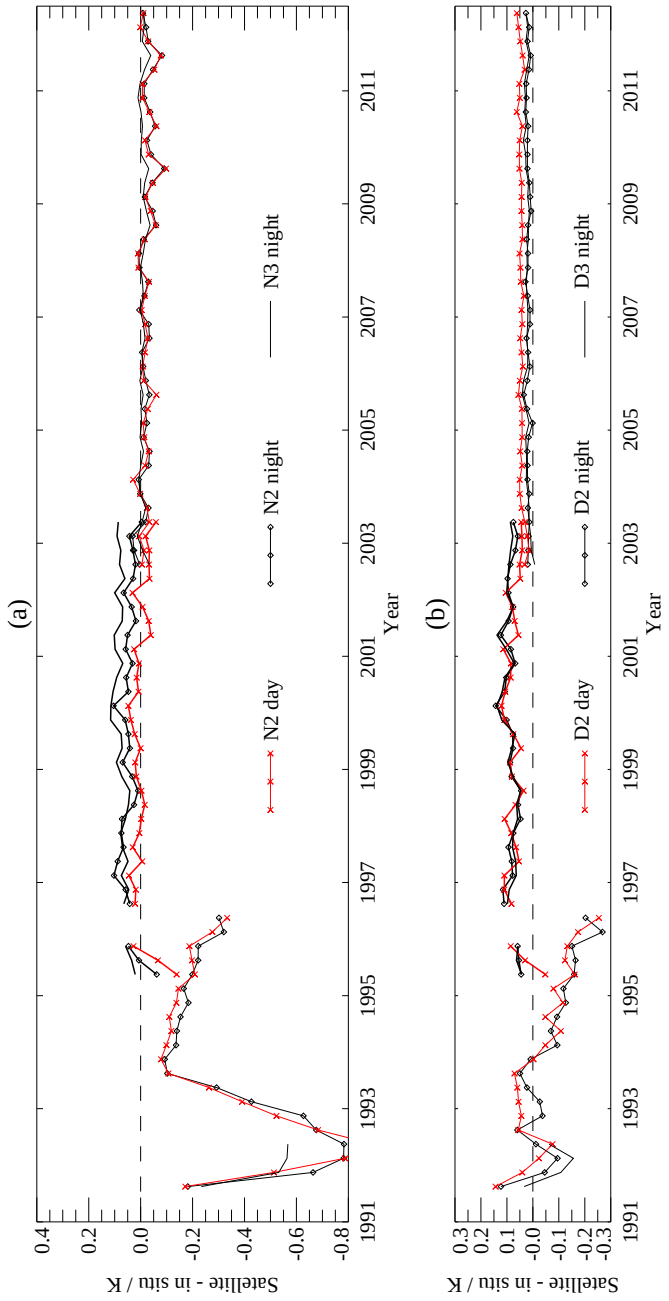

Figure 5.1: Satellite-based $SST_{0.2\,m}$ minus drifting buoy SST during operation of ATSR instruments. Time series shows global quarterly averages for ATSR-1 (up to mid-1996), ATSR-2 (mid-1995 to mid-2003), and AATSR (from mid-2002). (a) Nadir-only retrievals: solid with symbol N2; solid N3. (b) Dual-view retrievals: solid with symbol D2; solid D3. Red lines (with X symbol) indicate day-time data.

the dual-view SSTs, despite the dual-view retrieval coefficients being designed to be robust to this aerosol. Secondly, the early failure of the 3.7 µm channel means the "best" retrieval (dual-view three channel, D3) is only available for the first few months of operation of ATSR-1 (solid black line, Fig. 5.1b). Finally, ATSR-1 sensor instability (discussed later) particularly from 1994 onwards is reflected in SSTs not being stable with respect to drifters, culminating in a ~0.2 K discrepancy between ATSR-1 and ATSR-2 during the overlap between the two sensors in 1995.

In order for the ARC SSTs to meet the stability required for an SST climate data record, we must greatly reduce these inter-satellite discontinuities. The most direct option would be to examine the SST differences during the satellite overlap and adjust the SSTs from one sensor to agree with the other. Veal et al. (2013) take this approach for homogenizing the operational ATSR SSTs by subtracting a spatially varying mean bias field from the ATSR-2 data. This approach requires arbitrary assumptions as to how the SST differences observed during the overlap can be interpolated in space (to points observed only by one sensor) and extrapolated in time to adjust SSTs outside the overlap period. Here, instead of adjusting the SSTs to hide the discontinuities between sensors, we address the root of the discontinuities: inconsistencies between sensor calibration and/or deficiency in radiative transfer (RT) simulation.

Since the ARC SST retrievals are based on RT simulations, biases between simulations and observations will result in biased SST. (We remain agnostic as to whether the relative biases arise from sensor calibration, deficiency in RT simulation, or both — except in the case of the 12 µm channel of AATSR, as discussed below, where an instrument calibration problem is indicated.) As shown in Merchant and Le Borgne (2004), the retrieval biases can be written $\mathbf{a}^{\mathrm{T}}\varepsilon$ where the vector $\mathbf{a}$ lists the retrieval coefficients (weights) given to each channel in the SST retrieval algorithm, and ε lists the biases in the channels. The SST retrieval biases are typically larger than the RT biases because the coefficient vector has a magnitude larger than one, i.e. $|\mathbf{a}| > 1$. Moreover, the coefficients, $\mathbf{a}$, are banded by expected total column water vapour (TCWV) as described in Embury and Merchant (2012). The banding is necessary to minimise the regional retrieval biases which affect all coefficient-based retrieval schemes (Merchant et al., 2006b). Consequently, the spatial variability of the SST retrieval bias will combine variable biases in RT simulations.

For example, in the case of AATSR there is a calibration uncertainty in the 12 µm channel which appears to be biased cold by 0.2 K (G. Corlett, personal communication). This can be considered equivalent to the RT simulations being biased warm by 0.2 K for investigating the impact on SST retrieval. Figure 5.2 shows the effect of a constant 0.2 K bias in the 12 µm channel on the SST retrieval. The simulated retrieval bias varies as a function of total column water vapour (TCWV) and latitude, and is different for the various channel combinations used for ATSR SST retrievals. For instance, the dual-view night time (D3) retrieval is least affected as it gives more weight to the short wave 3.7 µm channel, while the dual-view day time (D2) retrieval is more strongly affected and displays a different dependence on TCWV or latitude. In practice the simulation bias will not be constant or restricted to a single channel and may vary as a function of TCWV or BT. Combining this with the effects of the

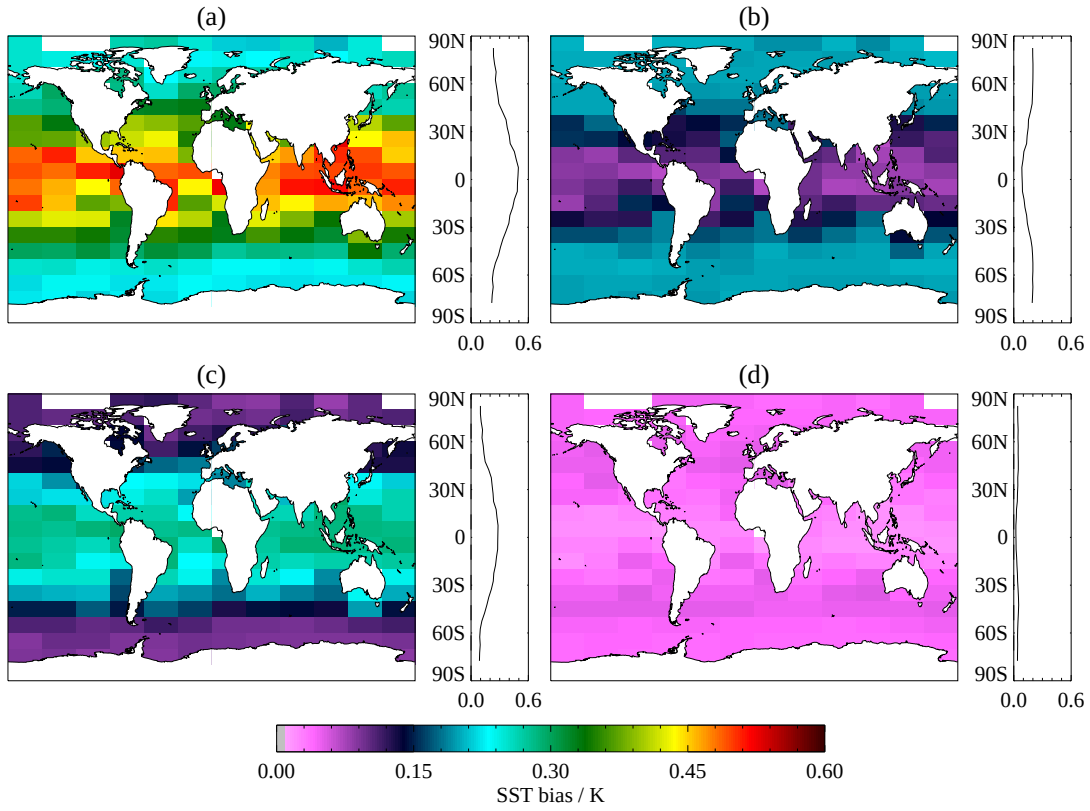

Figure 5.2: Effect on SST retrieval of an unaccounted for constant bias in brightness temperature (BT) calibration. Simulated SST biases due to a constant -0.2 K bias in 12 µm BT on (a) N2, (b) N3, (c) D2, and (d) D3 retrievals.

coefficients, the spatial variability of the SST retrieval will be more complex than the biases in the simulations.

Therefore we will address the discontinuities between sensors in terms of BT biases rather than SST biases in the paper. Section 5.2 discusses the calibration and characterisation of the ATSR instruments and how errors in them affect the satellite observations and the simulations which are the basis of the SST retrievals. Section 5.3 describes the inter-satellite harmonisation technique which is applied to the ATSR-2 to AATSR overlap in section 5.4 and the ATSR-1 to ATSR-2 overlap in section 5.5. Finally we check the stability of the resulting SST dataset by comparison to tropical moored buoys in section 5.6.

5.2 ATSR Calibration

5.2.1 Instrument design

The ATSR instruments were intended from the outset to allow accurate estimation of SST independently of in situ observations (Harries et al., 1983). The target calibration BT accuracy was 0.1 K over the dynamic range of interest and precision (radiometric noise) better than 0.08 K for the 3.7 µm channel and 0.05 for the 11 µm and 12 µm channels. Full details of the ATSR instrument design are

given in Edwards et al. (1990), Stricker et al. (1995), and Smith et al. (2012) for the ATSR-1, ATSR-2, and AATSR respectively. The key factors which contribute to the accurate calibration of ATSRs are briefly discussed below.

Firstly, the ATSRs used two on-board blackbody calibration targets, held at temperatures of 260 K and 300 K. These temperatures span the relatively narrow range of scene temperatures relevant for SST retrieval, which greatly reduced non-linearity effects in converting detector counts to BT. Both black bodies were observed once per scan, while the telescope was scanning the inside of the instrument in-between the nadir and forward views, ensuring that the same optical path was used for calibration and Earth views.

Secondly, the solid-state detectors that measured the infrared radiance were actively cooled to approximately 82 K to reduce instrument noise and avoid temperature dependent effects on the calibration. The instrument fore-optics were also cooled below ambient temperature reducing the impact of instrument self-emission and the variation of the self-emission due to temperature cycling. The instruments are enclosed and include baffles to minimise stray radiation affecting the detectors.

Finally, extensive pre-flight calibration and characterisation was performed to ensure the instruments operated within specification (see Mason, 1991; Smith et al., 1993; Smith, 1999; Smith et al., 2001). This includes accurate characterisation of the instrument spectral response functions (SRF) which is essential for performing the radiative transfer (RT) simulations needed for independent SST retrieval (Závody et al., 1995; Merchant et al., 1999; Embury et al., 2012a).

However, even with these efforts, both the ATSR-1 and AATSR instruments suffered complications with their 12 µm channels. Changes to the ATSR-1 detector temperature over time resulted in a shift in the 12 µm SRF, causing a warm bias of ~0.1 K in BT if the SRF shift were not accounted for. The AATSR 12 µm channel was affected by cold bias in BT of -0.2 K. While the cause is not fully determined, it is consistent with certain types of SRF measurement error (D. Smith, personal communication). The ARC SST retrievals were made immune to the error in the AATSR 12 µm channel by making all retrieval algorithms consistent with a reference retrieval using only the 3.7 µm and 11 µm channels, known as the D2* retrieval in Embury and Merchant (2012).

The following subsections discuss with reference to ATSR-1 how an incorrectly specified SRF affects both the calibration, i.e. the BTs reported by the instrument; and the characterisation, i.e. the BTs generated by the forward model.

5.2.2 Calibration error

Due to the way the first ATSR was interfaced with the ERS-1 platform, the instrument was subjected to greater warming than expected and the on-board cooler struggled to reach the target detector temperature of 82 K (Smith et al., 2012). In order to maximise the life of the cooler, the detector temperature (shown in Fig. 5.3) was initially set to 90 K and allowed to rise gradually, and exceeded 110 K at the end of routine operations in 1996. The increased detector temperature results both in

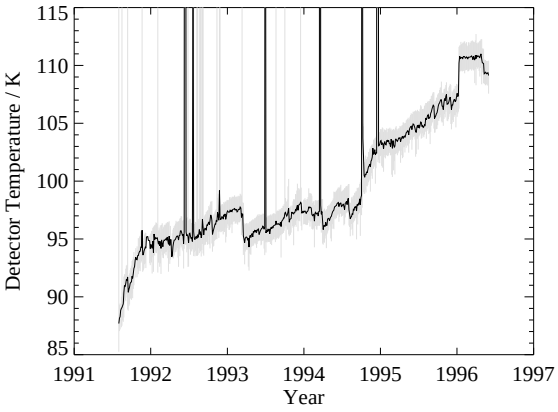

Figure 5.3: ATSR-1 detector temperature during operations. Solid line shows median detector temperature over three days, shaded grey area shows minimum and maximum values in period. Occasions where detector temperature goes off scale correspond to times when the cooler was switched off and the instrument warmed to ambient temperature.

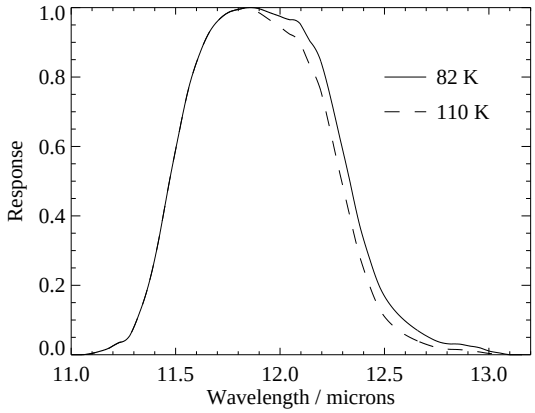

Figure 5.4: ATSR-1 12 µm spectral response function as measured during pre-flight calibration at 82 K (solid) and expected for end of life at 110 K (dashed).

increased noise in the observed BTs (Harris et al., 1995) and changes to the SRF of the 12 µm detector (Smith et al., 2012; Birks, 2008).

Birks (2008) examined in detail how the 12 µm detector was affected by higher operating temperatures and the resulting impact on the instrument BTs. In summary, as the temperature increases, so does the band gap of the photoconductive semiconductor which reduces the detector's sensitivity to long wavelength photons. Reducing the long-wave sensitivity changes the SRF, effectively shifting the long-wave cut-off to shorter wavelengths, resulting in a narrower SRF. Figure 5.4 shows the 12 µm SRF as it was measured in pre-flight calibration (82 K) and for a detector temperature typical of the end of mission (110 K) based on measurements of the flight-spare.

Birks (2008) investigated two main ways that the elevated detector temperature affects the calibration of the output BTs. The change to the SRF affects the radiance-to-BT conversion, and the detector

temperature also affects the detector non-linearity correction. Both of these effects are zero for scene temperatures close to either blackbody temperature, as illustrated using a simplified model of the ATSR calibration below.

The detectors return a signal, C, proportional to the radiance, R, reaching them. As the detectors observe both blackbodies and the scene through the same optical path the counts to radiance conversion can be written as:

$$w = \frac{C - C_1}{C_2 - C_1} \tag{5.1}$$

$$R(T) = (1 - w)R_1 + wR_2 \tag{5.2}$$

Where subscripts 1 and 2 refer to the two blackbody calibration targets. The temperature of the blackbody targets is known so the blackbody radiances, R_1 and R_2, can be calculated from the Planck function, B, and the detector SRF, ϕ, with:

$$R(T) = \int \phi(\lambda) B(\lambda, T) \mathrm{d}\lambda \tag{5.3}$$

This relation is inverted to calculate the BT corresponding to a given scene radiance, using a look-up table of numerical evaluations of R for different T.

Both the counts-to-radiance and radiance-to-BT conversions will be affected if the instrument SRF changes or was measured incorrectly. Writing the radiance for the erroneous SRF as $\tilde{R}$:

$$\tilde{R}(\tilde{T}) = (1 - w)\tilde{R}_1 + w\tilde{R}_2 \tag{5.4}$$

The error in the radiance, $\tilde{R} - R$, is the difference between Eq. 5.4 and 5.2 and therefore a linear function of w or the measured counts. Writing the BT error as:

$$\tilde{T} - T = \tilde{R}^{-1}\left(\frac{(R_2 - R(T))\,\tilde{R}_1 + (R(T) - R_1)\,\tilde{R}_2}{R_2 - R_1}\right) - T \tag{5.5}$$

Which is zero for $T = T_1$ or $T = T_2$ where the scene temperature matches either of the blackbody temperatures. Figure 5.5 shows the ATSR-1 12 µm channel at end-of-life (detector temperature = 110 K). Between the blackbody temperatures the maximum SRF error is ~5 mK. This is well within the ATSR accuracy specification, but nonetheless, as a change over time, is large enough to be significant for the stability of SST measurement for climate (discussed further below).

The detector non-linearity correction included in the full ATSR calibration model (Smith et al., 2012) is also affected by the detector temperature. Birks (2008) shows that this effect is comparable in magnitude to the SRF effect with the maximum combined calibration error reaching ~13 mK at the end of the ATSR-1 mission. Both effects can be accounted for using the correction from Birks (2008):

$$T = \tilde{T} + dT_m \frac{(T_2 - \tilde{T})(\tilde{T} - T_1)}{400}\left(1 + c_1(\tilde{T} - T_r)\right) \tag{5.6}$$

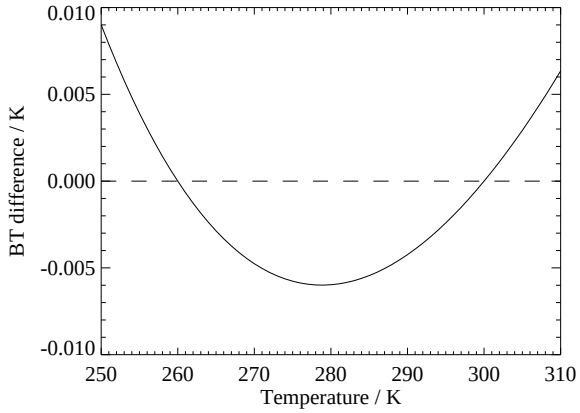

Figure 5.5: ATSR-1 12 µm calibration error due to using incorrect SRF at end-of-life.

Where dT_m is the maximum calibration error for detector temperature T_D given by:

$$dT_m = c_2(T_D - T_{D\text{ref}})^2 - c_3(T_D - T_{D\text{ref}}) + c_4 \tag{5.7}$$

With constants: $T_r = 280$, $T_{D\text{ref}} = 82$, $c_1 = 8.607 \times 10^{-3}$, $c_2 = 2.29718 \times 10^{-5}$, $c_3 = 2.27974 \times 10^{-4}$, and $c_4 = 1.4229 \times 10^{-3}$.

5.2.3 Characterisation error

The previous section showed that the ATSR-1 calibration error, even at end-of-life, is expected to be <0.015 K for the 12 µm BT. This does not necessarily capture the full impact on SST retrieval, since the changed SRF will also affect the RT simulations used to define the SST retrieval coefficients (Závody et al., 1995; Embury et al., 2012a). This occurs because the spectral radiance emitted by the Earth is not the same as the Planck curve emitted by a blackbody. The spectral radiance for a given frequency will be sensitive to the temperature over some range of height in the atmosphere or to the surface temperature, depending on the transmission of the atmosphere at that frequency. Changing the SRF changes the relative weighting of atmospheric and surface emissions contributing to the measured signal. The scene radiance observed by the instrument is:

$$R(T) = \int \phi(\lambda) L(\lambda) \mathrm{d}\lambda \tag{5.8}$$

where $L(\lambda)$ is the ToA radiance emitted by the Earth. The scene radiance is converted to an equivalent BT by inverting Eq. 5.3 as before.

Figure 5.6 shows simulations of the top-of-atmosphere BTs for the ATSR-1 12 µm channel using SRFs corresponding to 110 K and 82 K detector temperatures. The warmer detector temperature towards the end of life results in higher BTs if the corresponding modification to the SRF is not accounted for. This is due to the higher detector temperature reducing the long-wavelength response

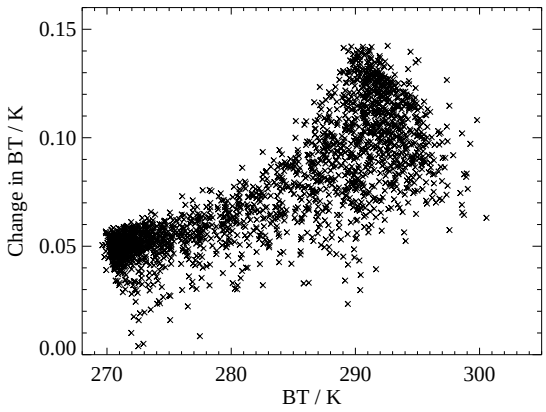

Figure 5.6: Change in simulated ATSR-1 12 µm BT due to altered SRF at end of life.

(see Fig. 5.4) such that the SRF gives greater weight to wavelengths where the atmosphere is relatively transparent. The magnitude of the BT impacts is 0.05 to 0.15 K. This is approximately 10 times larger than the calibration error, and the effect is strongly correlated with atmospheric water vapour.

In order to account for the change in SRF, the ARC SST retrieval coefficients are calculated for detector temperatures between 82.5 K and 110 K in steps of 2.5 K. This extends the SST retrieval coefficient lookup tables described in (Embury and Merchant, 2012) to include detector temperature as an extra dimension.

5.2.4 Practical notes

In this section, we have considered the errors associated with the increased ATSR-1 detector temperature in two parts: calibration and characterisation. The calibration error refers to the output of the instrument and covers the conversion from counts to radiance to BTs. Due to the design of the ATSR instruments the calibration error from the change in detector temperature is small $\lesssim$13 mK. The characterisation error covers the difference between what the instrument is measuring and what we expected it to measure. For the ATSR-1 instrument the pre-flight characterisation at a detector temperature of 82 K did not match the spectral response in flight when operating at higher detector temperatures. This resulted in the observed BTs being warmer than the simulated BTs by up to 0.15 K.

The distinction between calibration and characterisation errors is useful for physical based SST retrievals using forward modelling of the atmosphere (e.g. Závody et al., 1995; Merchant et al., 1999; Embury et al., 2012a) where an incorrect SRF can be considered part of the forward model error. However, for empirical based regressions such as Pathfinder (Casey et al., 2010) the distinction is less useful as the retrieval is based on regression relationships observed between BTs and in situ SST. No forward model is then required, so the characterisation error may be ignored, so long as it is the same for the BTs used for training the retrieval and the BTs to which the retrieval is applied.

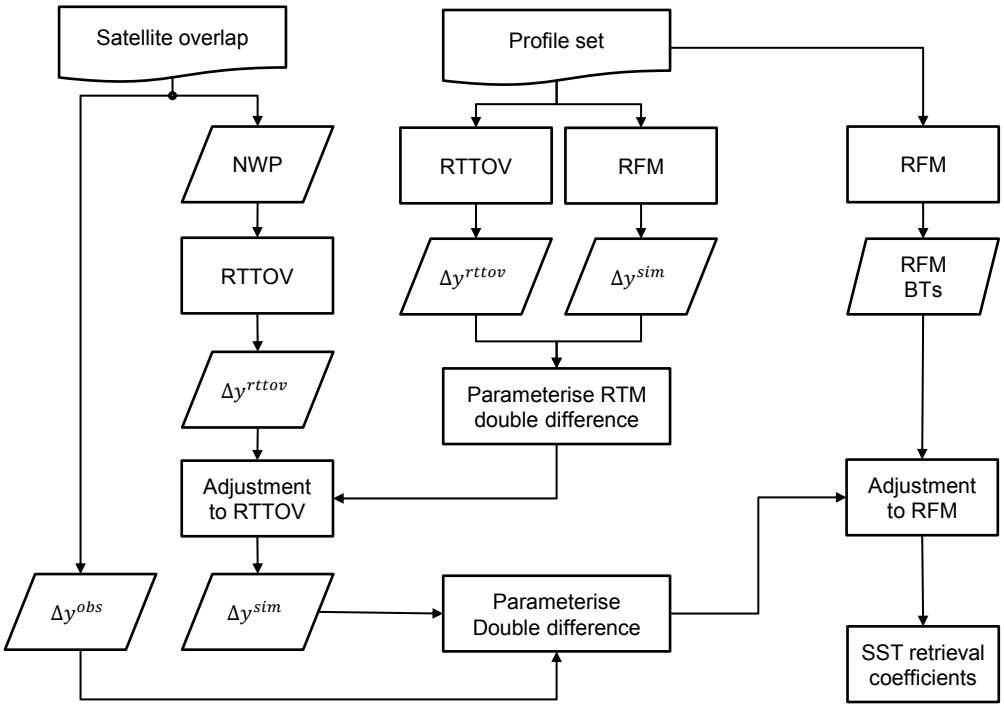

Figure 5.7: Flowchart of BT harmonisation process between two ATSR instruments. RTTOV is the fast RT model, RFM is the full line-by-line RT model used to generate SST retrieval coefficients, $\Delta\mathbf{y}$ is the BT difference between the two instruments.

5.3 Brightness Temperature Harmonisation

5.3.1 Overview

Our objective is an SST record which is consistent across sensors; i.e. coincident observations should result in the same SST retrieval. The approach we take is to harmonise the BTs in such a way that retrieved SSTs are consistent between sensors. This involves cross-calibrating the series of ATSR instruments using the ATSR-1/2 and ATSR-2/AATSR overlap periods. Previous approaches for sensor cross-calibration have been reported in Merchant et al. (2003), and Liang and Ignatov (2011). Here, cross-calibration involves taking proper account of real differences between sensors when specifying retrieval coefficients, and making adjustments for unattributable differences between sensors arising from calibration and/or characterisation errors. An overview of the process is shown in Fig. 5.7 and elaborated below.

For two sensors during an overlap period, we can extract observed differences for matched pairs of observations:

$$\Delta\mathbf{y}^{\text{obs}} = \mathbf{y}_1^{\text{obs}} - \mathbf{y}_2^{\text{obs}} \tag{5.9}$$

and we can also simulate the expected differences for these same pairs:

$$\Delta\mathbf{y}^{\text{sim}} = F(\mathbf{x}, \phi_1) - F(\mathbf{x}, \phi_2) \tag{5.10}$$

Here, the vector **y** lists the BTs in the 3.7, 11 and 12 µm channels, and subscripts 1, and 2 refer to the two overlapping sensors (i.e. could be ATSR-1 and ATSR-2, or ATSR-2 and AATSR). F is the simulation model which takes two main inputs: a state vector, **x**, representing the known atmospheric and surface information; and the SRF, ϕ, for the instrument. (The simulation model and observation matching procedure are discussed later.)

The observed BTs are subject to two error terms: radiometric noise and calibration error. The radiometric noise will average to insignificant levels over a sufficient number of observations, leaving the calibration error, which is expected to be small as explained in subsection 5.2.2

The simulated BTs are affected by several error terms. Firstly, we use NWP model data as the state vector input to the forward model and the NWP data will differ from the true state. The component that has the most significant effect on the simulated BTs is the NWP SST. Secondly, the forward model itself is subject to errors due to factors such as the spectroscopic data and the numerical approximations used. Finally the simulated BTs are sensitive to the instrument characterisation error discussed in subsection 5.2.3

For co-incident simulations, the NWP errors are common to both sets of simulated BTs and so they will cancel out in the simulated difference, $\Delta\mathbf{y}^{\mathrm{sim}}$. For matched (but not exactly co-incident) simulations, NWP-related effects on simulated differences may be assumed to be small and to cancel out on average. To the extent that SRFs are similar between the ATSR sensors, systematic forward model errors are also largely common to both sets of simulations and cancel out in the simulated difference. Systematic forward model errors may be present in differences although likely to be small: for instance, an error in spectroscopic data at a wavelength at the edge of one of the channels could affect one instrument more than another due to the differences in the SRFs.

Several error sources in the simulated differences therefore average out over many cases and/or are likely to be small. Significantly, however, any characterisation error in either of the SRFs will be particular to that instrument and will be systematically present in the simulated differences.

We can calculate the total effect of all the unknown calibration, characterisation, and forward model errors from the double differences $\Delta\mathbf{y}^{\mathrm{obs}} - \Delta\mathbf{y}^{\mathrm{sim}}$. If the double differences can be systematically characterised, this can then be used to account for the unexplained discrepancy between sensors in the un-harmonised ARC time series. This requires that we choose one of the four BT vectors to be adjusted (the four BT vectors being the observations for each sensor and the BT simulations used to define SST retrieval coefficients for each sensor). If the adjustment is applied to the BT simulations, this changes the SST retrieval coefficients from those found using unadjusted simulations, effectively bringing the simulations into agreement with observations. Alternatively, the adjustment could be applied to BT observations, without recalculating the SST retrieval coefficients.

From the previous section we expect the calibration errors to be smaller than the characterisation and forward model errors. Furthermore, it seems that the more recent instrument in each pair is better characterised than the older (with the exception of the AATSR 12 µm channel). In the light of these considerations, we choose to use the double difference characterisation to adjust the simulated

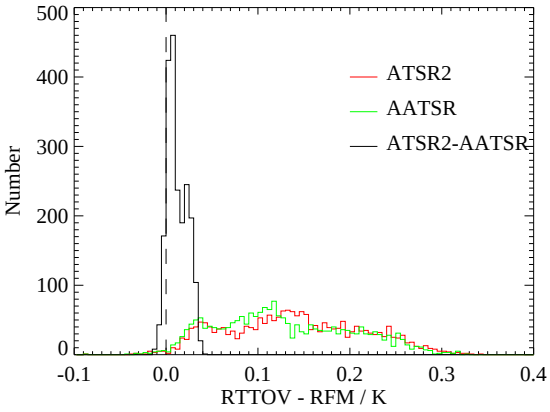

Figure 5.8: Histogram of RTTOV-RFM differences for 11 µm channel.

BTs (and thus the ARC SST retrieval coefficients) of the older instrument in each overlap pair. This means that all the SSTs are effectively harmonised to the BT calibration of the 3.7 and 11 µm channels of AATSR.

5.3.2 RTM double difference

One difficulty which arises is that the forward model used as the basis of the ARC SST retrieval coefficients is a full line-by-line model RFM (Dudhia, 2006), which is too slow to be practical to run for every pair of inter-sensor matchups. Instead, we use the fast forward model RTTOV 8.7 (Saunders and Brunel, 2005) to calculate the simulated differences and then account for the difference between the fast and full RT models. If we write the RTTOV simulated differences as:

$$\Delta\mathbf{y}^{\text{rttov}} = f(\mathbf{x},\phi_1) - f(\mathbf{x},\phi_2) \tag{5.11}$$

where f is the fast model approximation to the line-by-line model F. We need to use $\Delta\mathbf{y}^{\text{rttov}}$ to estimate $\Delta\mathbf{y}^{\text{sim}}$. Note that we do not have to improve the fast model, only the differences between sensors as simulated with the fast model.

Figure 5.8 shows a histogram of the RTTOV-RFM differences for 11 µm channel on ATSR-2 and AATSR using the profile set used to generate the ARC retrieval coefficients (Embury et al., 2012a). For the individual sensors there is a large spread in RTTOV-RFM differences spanning 0.0 to 0.3 K (with standard deviation of 0.08 K). However, when considering the double difference (the ATSR-2 RTTOV-RFM difference minus the AATSR RTTOV-RFM difference), the range is only 0.05 K and the standard deviation is 0.01 K. Furthermore, the remaining spread is strongly correlated with total column water vapour (TCWV) in the simulation, as shown in Fig. 5.9.

Therefore for ATSR-2/AATSR, we can estimate the $\Delta\mathbf{y}^{\text{sim}}$ consistent with the RFM model from the fast model RTTOV:

$$\Delta\mathbf{y}^{\text{sim}} = \Delta\mathbf{y}^{\text{rttov}} + \mathbf{a}_0 + w\mathbf{a}_1 \tag{5.12}$$

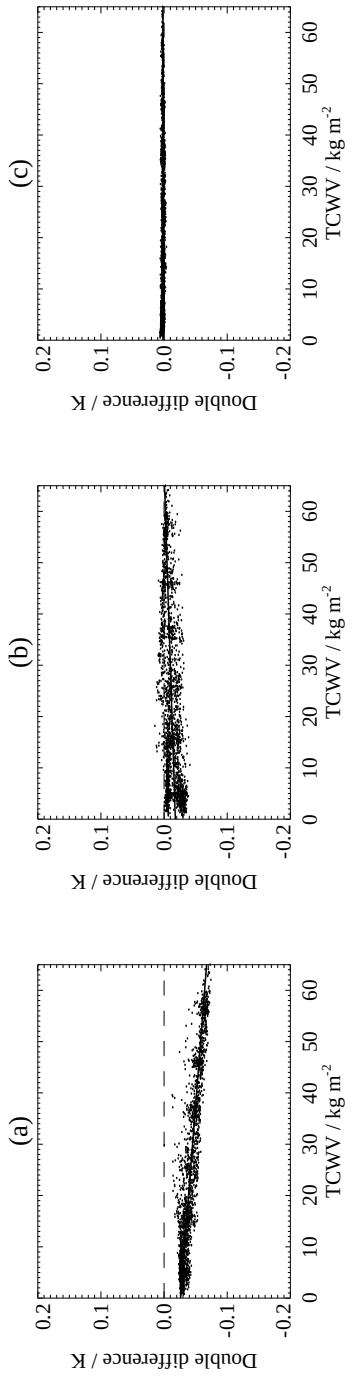

Figure 5.9: Double difference of RFM (ATSR-2 - AATSR) - RTTOV(ATSR-2 - AATSR) as a function of total column water vapour (TCWV). (a) 3.7 µm nadir, (b) 11 µm nadir, (c) 12 µm nadir

where w is TCWV, and $\mathbf{a}_0$ and $\mathbf{a}_1$ are coefficients found from a linear regression to the difference $\Delta\mathbf{y}^{\mathrm{rfm}} - \Delta\mathbf{y}^{\mathrm{rttov}}$ calculated over the reference profile set.

5.3.3 Match-up and time difference adjustment

Generating inter-sensor matches between the three instruments is simplified by the fact that the satellites operated in the same orbital plane during the overlap periods to allow inter-satellite comparisons. This tandem operation brings significant benefits when exploiting co-located matches, as both sensors observed the scenes with the same viewing geometry etc. However, flying in the same orbital plane means there must be a temporal separation between the satellites. ERS-2 (on which flew ATSR-2) was placed in an orbit 24 hours behind ERS-1 (ATSR-1), while Envisat (AATSR) was 30 minutes ahead of ERS-2.

Due to the small time separation between Envisat and ERS-2 it is easy to identify large numbers of AATSR to ATSR-2 matchups which are cloud free for both sensors. However, the 30 minutes separation is sufficient for the diurnal cycle in SST to have a significant effect on $\Delta\mathbf{y}^{\mathrm{obs}}$, particularly under conditions of low wind speed where the diurnal warming can be 0.3 K/hr (Embury et al., 2012b). The ARC SST products include an adjustment to account for the different overpasses of ERS-2 and Envisat based on the Fairall et al. (1996) skin and Kantha and Clayson (1994) stratification models (see Embury et al., 2012b; Horrocks et al., 2003b, for details) which can be used to calculate $\Delta\mathbf{y}^{\mathrm{sim}}$ accounting for the appropriate 30 minute change in SST.

For ATSR-1 and ATSR-2 the 24 hour separation between observations leads to a substantial reduction in the number of match-up which are clear-sky for both sensors. We find a sufficient number at low and mid-latitudes, but at high-latitudes there are hardly any 24-hour matches which remain clear for both observations, due to the higher frequency of cloudy conditions at high-latitudes. However, polewards of 55 degrees the sensor swaths overlap and it is possible to get matches separated by approximately 31 minutes. Unlike the 24-hour matches, these high-latitude matches do not have common viewing geometry. This is accounted for in the double differences as $\Delta\mathbf{y}^{\mathrm{sim}}$ is calculated using the appropriate zenith angles.

5.4 ATSR-2 to AATSR

5.4.1 Inter-satellite differences

Figure 5.10 shows $\Delta\mathbf{y}^{\mathrm{obs}}$ and $\Delta\mathbf{y}^{\mathrm{sim}}$ calculated for ATSR-2 - AATSR as a function of TCWV for the three channels and two views. For the 3.7 µm channel (panels (a) and (d)), both simulation and observation indicate that AATSR is returning warmer BTs than ATSR-2. The observed differences (error bars with no symbol) in the nadir view are smaller than expected from simulation: from $\Delta\mathbf{y}^{\mathrm{obs}}$ we see that ATSR-2 and AATSR observations within 0.1 K of each other, while $\Delta\mathbf{y}^{\mathrm{sim}}$ (the difference expected from the SRF differences, shown in errors bars with star symbols) goes more negative than -0.2 K at high

	$\mathbf{a}_0^{\text{nad}}$	$\mathbf{a}_1^{\text{nad}}$	$\mathbf{a}_0^{\text{fwd}}$	$\mathbf{a}_1^{\text{fwd}}$
3.7 µm	0.025076	0.001606	0.015709	0.002293
11 µm	0.008138	0.000640	0.011338	0.000664
12 µm	0.081881	0.003347	0.087100	0.003911

Table 5.1: Coefficients for adjustment of ATSR-2 BTs to be more consistent with AATSR BTs after accounting for reported spectral response function differences.

TCWV. The forward view shows slightly larger differences than the nadir, but both $\Delta\mathbf{y}^{\text{obs}}$ and $\Delta\mathbf{y}^{\text{sim}}$ share the same behaviour. For the 11 µm channel the simulated and observed differences are in much closer agreement with the difference between simulated and observed differences less than 0.05 K. The 12 µm channel shows the greatest disagreement—simulations indicate that the two sensors should be returning almost identical BTs because of very similar SRFs, while in practice ATSR-2 observations are 0.1 to 0.3 K warmer than AATSR observations.

Figure 5.11 shows the double difference $\Delta\mathbf{y}^{\text{obs}} - \Delta\mathbf{y}^{\text{sim}}$ between the two instruments. The double differences all have a linear dependence on water vapour and only a minimal difference between the nadir and forward views. The double difference for the 12 µm channel shows that AATSR 12 µm BTs have a 0.2 K cool bias on average, as reported, although the double difference makes clear a strong dependence on TCWV. The double difference for the 3.7 µm channel shows a larger discrepancy than the 11 µm channel, which explains why the N3 and D3 SST retrievals (which use the 3.7 µm channel) show a larger discrepancy than N2 and D2 SSTs between ATSR-2 and AATSR in the unharmonised SST time series in Fig. 5.1.

5.4.2 Corrections for ATSR-2

The double differences from Fig. 5.11 were fitted with a linear function of TCWV:

$$\Delta\mathbf{y}^{\text{obs}} - \Delta\mathbf{y}^{\text{sim}} = \mathbf{a}_0 + w\mathbf{a}_1 \tag{5.13}$$

The coefficients $\mathbf{a}_0$ and $\mathbf{a}_1$ are shown in Table 5.1.

Adjustments for the 3.7 and 11 µm channels can be used to harmonise the ATSR-2 and AATSR SSTs either by applying the adjustments to the observed BTs and using the existing SST retrieval coefficients, or by applying the adjustments to the simulated BTs and calculating new SST retrieval coefficients. (The 12 µm adjustment is not used. For this channel on AATSR, the calibration error was known to be significant prior to this work on harmonisation, and thus the ARC ATSR-2 and AATSR retrieval coefficients were designed so as to avoid dependence on the calibration of the 12 µm channel, as described in Embury and Merchant (2012)) Although the harmonisation adjustments can be applied to either the simulations or observations there are two reasons why we choose to apply the harmonisation adjustments to the simulated BTs.

Firstly, by applying the adjustments to the simulations we are implicitly assuming that the simulations are in error and the observations are correct. This is consistent with the relative impacts

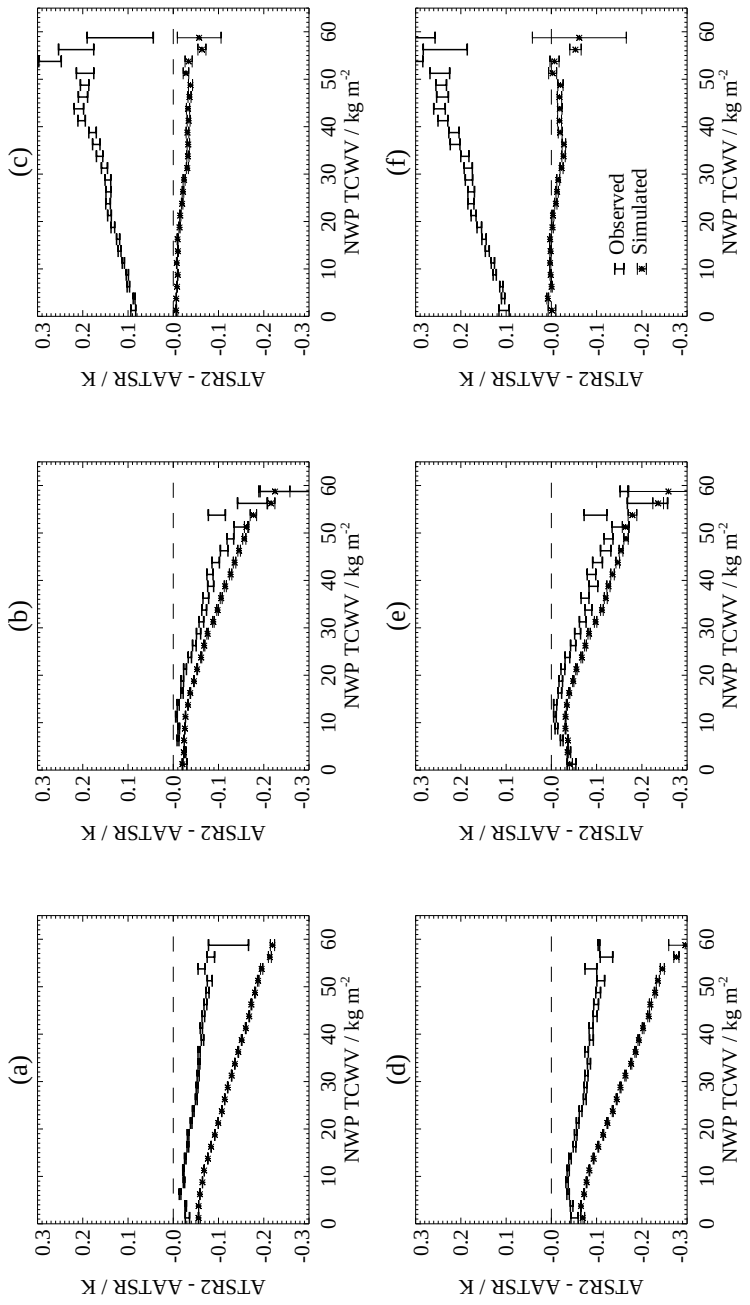

Figure 5.10: Observed (black) and simulated (black with symbol) ATSR-2 - AATSR BT differences as a function of TCWV. (a) 3.7 µm nadir, (b) 11 µm nadir, (c) 12 µm nadir, (d) 3.7 µm forward, (e) 11 µm forward, (f) 12 µm forward.

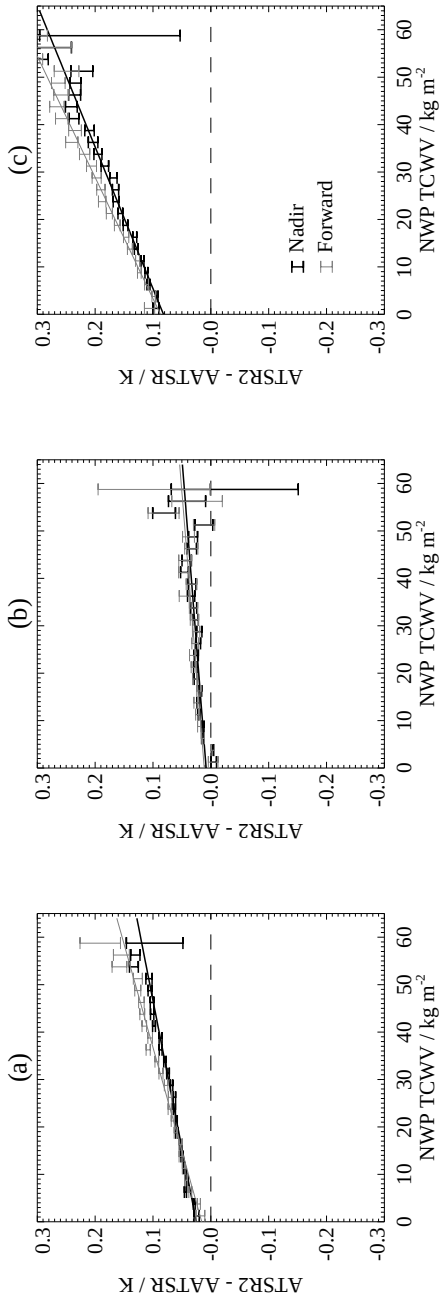

Figure 5.11: Double difference ($\Delta\mathbf{y}^{\mathrm{obs}} - \Delta\mathbf{y}^{\mathrm{sim}}$) between ATSR-2 and AATSR for nadir view (black) and forward view (grey). (a) 3.7 µm channel, (b) 11 µm channel, (c) 12 µm channel.

of known calibration and characterisation errors presented in section 5.2, where the increased detector temperature on ATSR-1 resulted in a smaller calibration error (error in the observations) than characterisation error (error in the simulations).

Secondly, applying the adjustments to the simulations is more convenient from a practical point of view. The $\Delta\mathbf{y}^{obs} - \Delta\mathbf{y}^{sim}$ double difference is tied to the specific forward model (RFM) which was used to generate the retrieval coefficients and the assumption of clear-sky over ocean. The ARC processing chain also includes cloud detection (Merchant et al., 2005; Mackie et al., 2010) and a Saharan dust index (Good et al., 2012) which were developed prior to the BT harmonisation and require the unmodified observations.

5.4.3 AATSR 12 µm effect

While the double difference for the 12 µm channel is not used for harmonisation as mentioned above, it does provide some insight into the behaviour of the channel on AATSR. For instance, the direction of bias and dependence on TCWV is consistent with the AATSR 12 µm channel being more sensitive to longer-wavelength radiation than expected from the measured SRF. This is one of the proposed explanations for the observed bias (O'Hara et al., 2013). Moreover, the coefficients from Table 5.1 could be used as an adjustment to RT model outputs for clear-sky ocean scenes, instead of using a constant 0.2 K as was done for the ARC reprocessing (Embury and Merchant, 2012), which may further improve the efficiency of ARC cloud detection.

5.5 ATSR-1 to ATSR-2

5.5.1 Inter-satellite differences

Figure 5.12 shows $\Delta\mathbf{y}^{obs}$ and $\Delta\mathbf{y}^{sim}$ calculated for ATSR-1 - ATSR2 as a function of TCWV for the two channels available during the instrument overlap. The differences show a more complex dependence on TCWV that the ATSR-2 to AATSR ones shown in Fig. 5.10. The fact that the same dependencies on TCWV are seen in both the simulations and observations indicates that this behaviour is explained by the interaction of SRF differences and the IR spectrum in different TCWV conditions (rather than merely being sampling effects, for example). Part of this will be due to the tropical and mid-latitude matches being separated by 24 hours and the change in atmospheric conditions over that time, which are captured in the NWP used to drive the forward model.

The double differences for ATSR-1 minus ATSR-2 are shown in Fig. 5.13. The magnitude of the double differences is comparable to that seen in Fig. 5.11 for the ATSR-2 to AATSR overlap and both nadir and forward views are consistent; however the double differences for ATSR-1 minus ATSR-2 are not a linear function of TCWV. The magnitude of the double difference for the 11 µm channel is smaller, mostly less than ~0.1 K, compared to the 12 µm channel which is up to ~0.2 K.

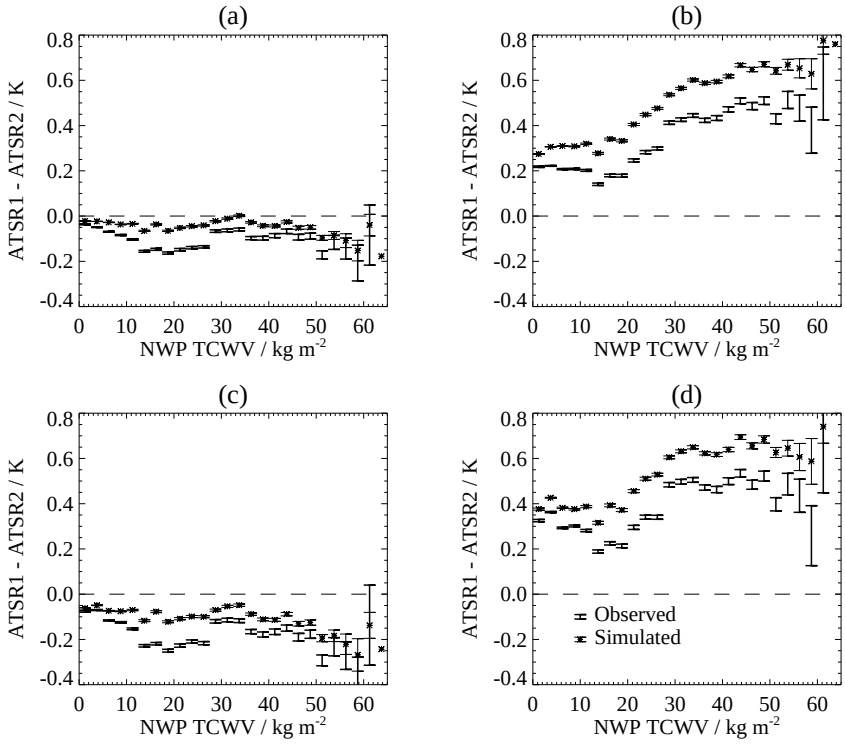

Figure 5.12: Observed (black) and simulated (black with symbol) ATSR-1 - ATSR-2 BT differences as a function of TCWV. (a) 11 µm nadir, (b) 12 µm nadir, (c) 11 µm forward, (d) 12 µm forward.

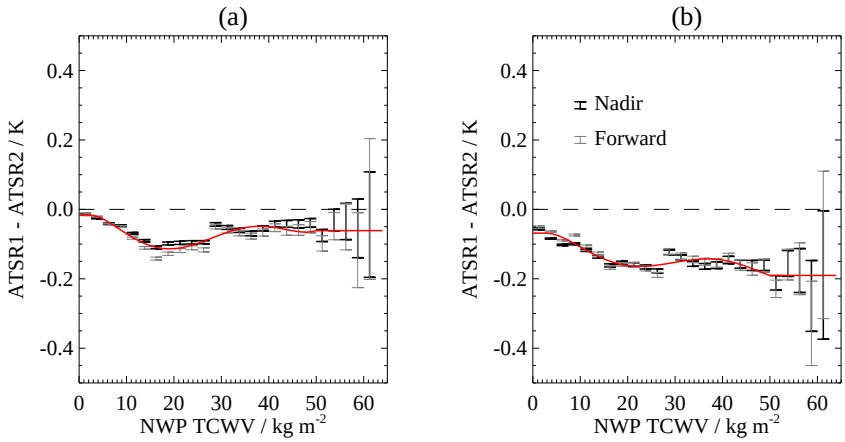

Figure 5.13: Double difference ($\Delta\mathbf{y}^{obs} - \Delta\mathbf{y}^{sim}$)) between ATSR-1 and ATSR-2 for nadir view (black) and forward view (grey). (a) 11 µm channel, (b) 12 µm channel.

	11 µm	12 µm
$\mathbf{a}_0$	-2.30773×10^{-2}	-7.87478×10^{-2}
$\mathbf{a}_1$	0.79284×10^{-3}	8.51432×10^{-3}
$\mathbf{a}_2$	-2.35391×10^{-3}	-1.98255×10^{-3}
$\mathbf{a}_3$	1.38492×10^{-4}	1.03176×10^{-4}
$\mathbf{a}_4$	-3.03028×10^{-6}	-2.07218×10^{-6}
$\mathbf{a}_5$	2.26715×10^{-8}	1.43136×10^{-8}

Table 5.2: Coefficients for adjustment of ATSR-1 simulations from overlap with ATSR-2.

5.5.2 Corrections for ATSR-1

The double differences from Fig. 5.13 were fitted with a polynomial function of TCWV (w):

$$\Delta\mathbf{y}^{\text{obs}} - \Delta\mathbf{y}^{\text{sim}} = \sum_{i=0}^{5} w^i \mathbf{a}_i \tag{5.14}$$

In order to constrain the fit at large TCWV where the standard errors are larger, the maximum TCWV used is 52.5 kg m^{-2}, and above this the correction reverts to a constant. For ATSR-1/2, there is not a clear distinction between the nadir and forward-view double differences, so a single correction per channel is produced. The coefficients are shown in Table 5.2.

5.5.3 Inter-algorithm adjustment

With ATSR-1 there is one more obstacle before harmonised SST retrieval coefficients can be generated –inter-algorithm consistency for ATSR-2/AATSR relied on the availability of the 3.7 µm channel (Embury and Merchant, 2012). As ATSR-1's 3.7 µm channel failed in May 1992, there are only seven months of data available to perform the inter-algorithm adjustment at the beginning of ATSR-1's life. This period spanned a rapid increase in detector temperature, from 88 to 95 K, and the heaviest loadings of of stratospheric aerosol from the Mount Pinatubo eruption (Lambert et al., 1997).

While the dual-view SST retrievals are designed to be stratospheric-aerosol robust (Merchant et al., 1999), the nadir-view retrievals are not, so any inter-algorithm adjustment made using the initial six months of data would only apply to the dual-view D2 and D3 retrievals. Furthermore, the 3.7 µm channel central to the inter-algorithm adjustment has not been harmonised with ATSR-2 because it is not available during the overlap period. As such is it not reasonable to extrapolate from the initial six month period to the rest of the ATSR-1 mission.

The only option to ensure that ATSR-1's RT-based coefficients are consistent across the different types of SST algorithm (i.e., when using different channel combinations) is to use the overlap with ATSR-2 to provide the inter-algorithm adjustments. (For an account of inter-algorithm consistency adjustments for ATSR-2 and AATSR, see Embury and Merchant (2012)). Such adjustment is only possible for the two-channel retrievals (N2 and D2).

Figure 5.14 shows the inter-algorithm adjustments calculated at start and end of ATSR-1 operations. Because the ATSR-1/2 overlap period is near the end of the ATSR-1 mission, the inter-algorithm

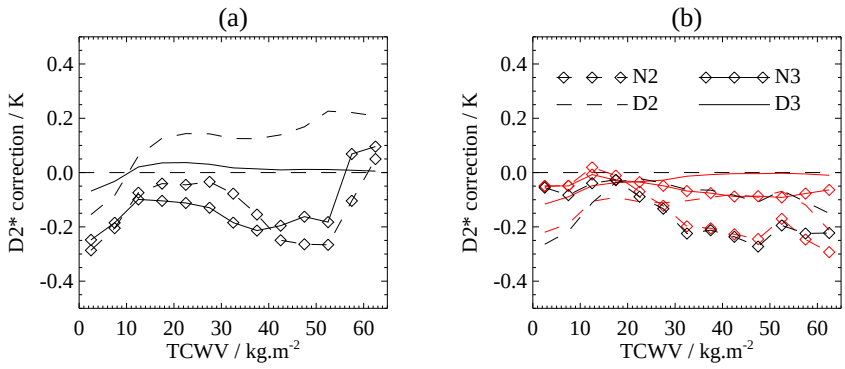

Figure 5.14: Inter-algorithm adjustments for ATSR-1: (a) relative to ATSR-1 D2* at start of mission, detector temperature ~95 K; (b) relative to ATSR-2 D2* at end of mission, detector temperature ~105 K. Solid line: D3; dashed line: D2; solid with diamonds: N3; dashed with diamonds: N2; red lines show ATSR-2 inter-algorithm adjustment for reference.

Retrieval	Start-of-life adjustment	End-of-life adjustment	Choice
N2	Degraded by aerosol	Available	Use end-of-life
N3	Degraded by aerosol	N/A	Use start-of-life
D2	Available	Available	Interpolate
D3	Available	N/A	Use start-of-life

Table 5.3: Inter-algorithm adjustments available for ATSR-1 at start and end of mission.

adjustment is correct only for detector temperatures around 105 K (panel (b)). For this time, the N2 and D2 adjustments are consistent with those calculated for ATSR-2 (see Fig. 12a in Embury and Merchant, 2012, also shown here in Fig. 5.14b in red)), should be the case given that the BTs have been harmonised. However, the adjustments need to be different at the start of the mission for two reasons. Firstly, at the start of the mission, the ATSR-1 3.7 µm channel is available (and the ATSR-2 D2* retrieval is of course not available). The coefficients for N2, D2, N3 and D3 are therefore referenced to the ATSR-1 D2* at the start of mission, rather than ATSR-2 D2* at the end. The 3.7 µm channel could not be harmonised with ATSR-2 equivalent, because it failed in May 1992, and this means that the D2* retrievals of the two instruments are not necessarily consistent. The other issue that may affect the adjustments is the stratospheric aerosol present at the start of life. The ATSR-1 D2* is designed to be robust to stratospheric aerosol, based on our knowledge of the properties of the aerosol and its effect on radiative transfer. But any residual sensitivity to aerosol in the D2* retrieval will be propagated, via the inter-algorithm adjustments, to all types of retrieval.

Table 5.3 shows the inter-algorithm adjustments available for ATSR-1 and which are used in the ARC products. For the N2 retrieval we use the end-of-mission adjustment as the start-of-mission adjustment will be degraded by the presence of stratospheric aerosol. For the N3 and D3 retrieval we the start-of-mission adjustments due to the failure of the 3.7 µm channel. For the D2 retrieval both start-of-mission and end-of-mission adjustments are available, so we interpolate between the two based on the detector temperature.

5.6 Stability Assessment

5.6.1 Rationale

The BT harmonisation adjustments from the previous two sections were used to adjust RFM-simulated BTs for ATSR-2 and ATSR-1 in such a way as to make them more consistent with AATSR. Having done this, we generated the final (harmonised) ARC SST retrieval coefficients following the techniques from Embury and Merchant (2012). SSTs using these harmonised coefficients should be more stable than the unharmonised ARC SSTs in Fig. 5.1. This expectation is assessed in this section.

The evolution of the harmonised ARC SST bias relative to in situ drifters is shown in Fig. 5.15. Compared to the unharmonised time series (Fig. 5.1), we see that the ~0.1 K discrepancies between instruments have been removed and the ATSR-1 biases appear more stable with time, albeit ATSR-1 SSTs are still less stable than the later data.

At the level of stability seen in Fig. 5.15, it is important to consider the evolution of the drifter network over the period. At the start of the time series there are approximately 5,000 match-ups per year between satellite and drifter observations, while in the AATSR period there are 60,000 or more match-ups per year. The geographical distribution of the matches is different, with significant data voids in the earlier periods. As such, the satellite-drifter differences in the early part of the record suffer from poorer sampling and may be more affected by biases from individual 'bad' drifters. Not necessarily all of the relative instability of the ATSR-1 in comparison to drifters arises from the problems of channel failure, detector temperature rise and stratospheric aerosol: some may be associated with the drifting buoy SSTs to which the satellite SSTs are compared in Fig. 5.15.

In order to provide an alternative assessment of the stability of the harmonised ARC SST products, we use the Global Tropical Moored Buoy Array (GTMBA) as described in McPhaden et al. (2010). The array uses Autonomous Temperature Line Acquisition System (ATLAS) moorings which provide SST to greater accuracy and precision than drifters. Moorings are serviced at six-month or yearly intervals to ensure continuity of data and recover data recorded at high temporal resolution. The oldest component of the GTMBA, the Tropical Atmosphere Ocean (TAO) array (McPhaden, 1995), was deployed in the tropical Pacific between 1985 and 1994. We consider that GTMBA data provide the most stable in situ source for evaluating the stability of the ARC SSTs, although the stability assessment is then limited to tropical latitudes.

For context we also present an equivalent comparison of AVHRR Pathfinder (Casey et al., 2010) against GTMBA. The Pathfinder SST retrieval is based on an empirical regression to in situ drifters (and ships data for earlier years). It may therefore be expected that Pathfinder SSTs should automatically be stable (at least, as stable as the in situ measurements) without particular steps to harmonise between sensors. Comparison with Pathfinder SST is therefore is a useful gauge of how successful the harmonisation of ARC SSTs has been. While the Pathfinder SSTs are tied to in situ measurements, GTMBA data are not used in the Pathfinder regression, so the comparison is (as with ARC) against

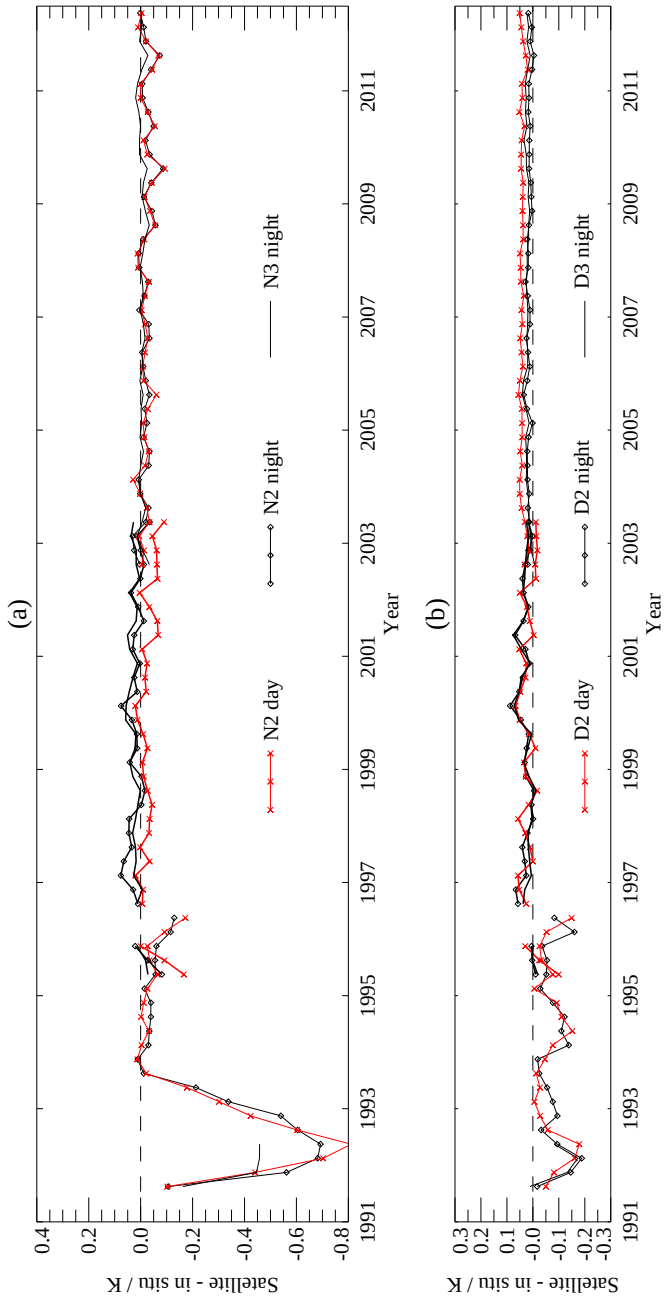

Figure 5.15: ARC $SST_{0.2m}$ minus drifting buoy SST. ATSR-1 data used up to mid-1996, ATSR-2 from mid-1995 to mid-2003, and AATSR from mid-2002 onwards. (a) Nadir-only retrievals: solid with symbol N2; solid N3. (b) Dual-view retrievals: solid with symbol D2; solid D3. Red lines (with X symbol) indicate day-time data.

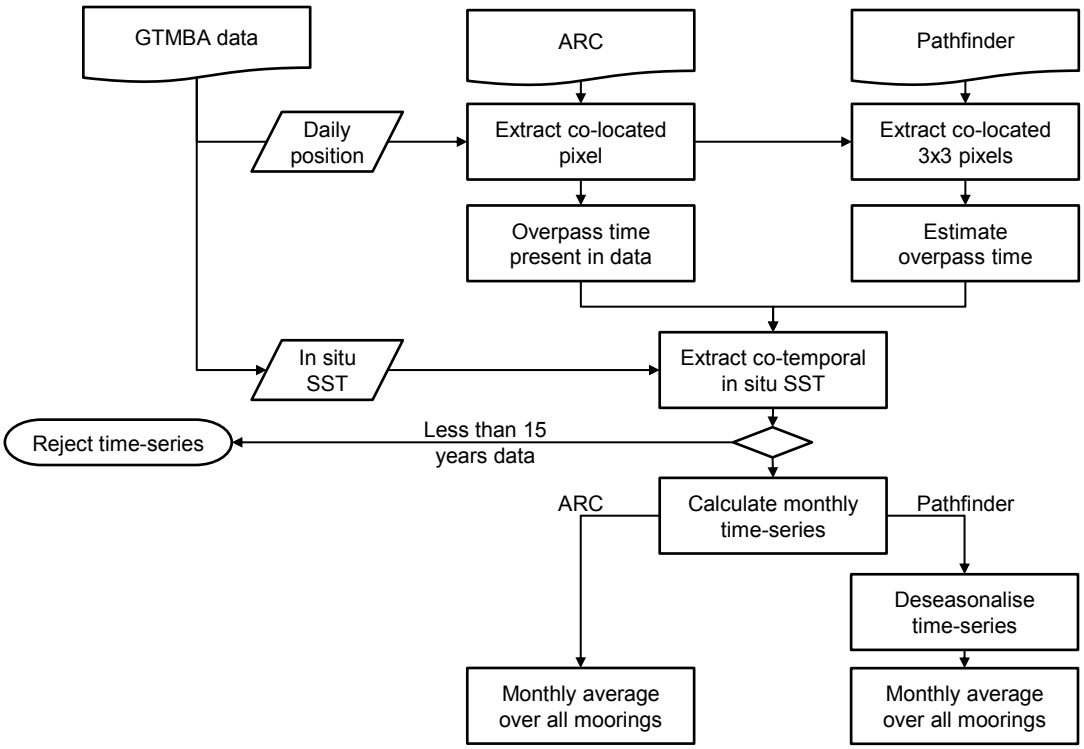

Figure 5.16: Flow chart of match-up process between GTMBA and satellite datasets.

independent data.

5.6.2 Match-up process

For the match-up process we used the following data sets: ARC SST v1.1.1 daily 0.1° gridded product (available from the Natural Environment Research Council Earth Observation Data Centre—see acknowledgements for details); the AVHRR Pathfinder v5.2 daily $\frac{1}{24}$° (approx. 4km) gridded product; and GTMBA high resolution delayed-mode files from the Pacific Marine Environment Laboratory (PMEL) website, which provides up to 10 minute temporal resolution. The match-up process between the datasets is outlined in Fig. 5.16. For each buoy we use the daily position files to identify the co-located satellite pixels and then the high resolution in situ data co-temporal with the satellite observations. This approach ensures that both the temporal and spatial separation between satellite and in situ is minimised.

The ARC SST products include three estimates: SST_{skin} at satellite overpass time; and $SST_{0.2m}$ and $SST_{1.0m}$ at a standardised time. The standardised time is 10:30 local equatorial crossing time (LECT) (corresponding to the ERS-1 and ERS-2 ascending node) and is used to avoid aliasing the diurnal cycle into the ARC products (Embury et al., 2012b; Merchant et al., 2012). For this comparison against moored buoys we used $SST_{1.0m}$ (since this depth best corresponds to the uppermost measurement on ATLAS moorings) and extract the single 0.1° cell encompassing the buoy location.

The Pathfinder dataset provides SST consistent with drifting buoys (and/or ships) at the satellite

overpass time. The LECT of the platforms carrying AVHRR instruments drifts over time (Jin and Treadon, 2003). Unfortunately, the current v5.2 Pathfinder files do not include the time of observation. Therefore we estimated the AVHRR observation time using the method of Ignatov et al. (2004) which parameterises the LECT of each satellite such that the overpass time for any point can be calculated given its longitude and latitude. This method will only be accurate to about an hour, as we also do not have the position of the data within the AVHRR swath and must assume it is centre swath. Due to the higher spatial resolution of the Pathfinder dataset we extract the 3×3 pixels centred on the buoy to match the ARC resolution approximately and use the mean of the highest quality (quality_level=5) pixels present.

The time series is then generated as follows:

1. Identify GTMBA moorings with at least 15 years of satellite matchups
2. For each mooring:
   (a) Calculate monthly time series of median satellite-buoy difference
   (b) (Found to be necessary for AVHRR only) Deseasonalise time series:
      i. Calculate multi-year average of monthly time series excluding the Pinatubo-affected years (2.a)
      ii. Subtract the multi-year average (2.b.i) from monthly series(2.a)
3. Calculate the monthly average across all moorings
4. Find least squares fit to time series of average differences

The constraint on the length of the mooring record is necessary to prevent changes in the in situ record affecting the comparison. In practice this limits the comparison to the TAO/PIRATA array in the Pacific.

The Pathfinder data is deseasonalised to prevent seasonal and regional biases being aliased into the composite time series and giving a pessimistic impression of Pathfinder SST stability. As the deseasonalisation is done on a per-mooring basis, any regional variation in bias is removed as part of the deseasonalisation for that particular buoy. For Pathfinder, the median satellite-in situ SST (before deseasonalisation) varied from -0.68 to -0.04 K depending on the mooring (using TAO/PIRATA moorings with at least 15 years of data and excluding data prior to 1995 to avoid effects of stratospheric aerosol from the Mount Pinatubo eruption). The robust standard deviation of the median satellite-mooring differences is 0.10 K. For comparison the median satellite-mooring differences for the ARC SSTs varied from -0.06 to 0.10 K with a robust standard deviation of 0.03 K.

5.6.3 Stability Assessment

The time series of monthly mean discrepancy between satellite and GTMBA is shown in Fig. 5.17. The ARC SST data are highly stable relative to GTMBA after 1995, but there is a drift for earlier data from

Product	Time of Day	95% confidence interval (mk year^{-1})
ARC v1.1.1	Day	-0.35 < trend < 2.03
ARC v1.1.1	Night	-0.13 < trend < 1.68
Pathfinder v5.2	Day	0.58 < trend < 5.59
Pathfinder v5.2	Night	2.59 < trend < 7.73

Table 5.4: Confidence intervals for estimated satellite - GTMBA trends from 1995 onwards estimated using ordinary least squares regression.

Product	Time of Day	95% confidence interval (mk year^{-1})
ARC v1.1.1	Day	1.60 < trend < 51.5
ARC v1.1.1	Night	-10.2 < trend < 36.3
Pathfinder v5.2	Day	24.8 < trend < 247
Pathfinder v5.2	Night	-20.8 < trend < 198

Table 5.5: Confidence intervals for estimated satellite - GTMBA trends prior to 1995 estimated using ordinary least squares regression.

ATSR-1 with ARC $SST_{0.2m}$ approximately 0.1 K cooler than GMTBA at the start of the record in 1991. The Pathfinder-GTMBA time series also shows different behaviour for pre and post-1995 data. Prior to 1995 the Pathfinder data is more variable with an excursion exceeding 1 K in magnitude corresponding to the Pinatubo eruption in 1991. After 1995 the Pathfinder data become substantially more stable, which corresponds to the switch from the AVHRR-11 to AVHRR-14 sensor.

Linear trends were fitted to data pre- and post-1995 as both ARC and Pathfinder show a change in behaviour at that point. Table 5.4 shows the trends estimated for the later period and Table 5.5 shows the trends for the earlier period. There is no significant trend in the ARC data after 1995 and the confidence intervals of -0.4 to 2.0 mK year^{-1} (daytime) and -0.1 to 1.7 mK year^{-1} (nighttime) are well within the target stability of 5 mK year^{-1}.

The simple trend analysis shown here uses an ordinary least squares regression which ignores autocorrelation error. As a result the reported trends may be over estimates of the true trend while the confidence interval may be under-estimated. If there is autocorrelation in the data, there is therefore a risk of type I error (i.e. incorrectly determining that a significant trend exists when it does not). However, the ARC-GTMBA time series in Fig. 5.17 does not appear to be autocorrelated after 1995. The trend estimates after 1995 are not significantly different from zero and very likely to be 2 mK year^{-1} or less in magnitude. The Pathfinder-GTMBA time yields a significant non-zero trend using this analysis. However, there is inter-annual variability evident in the Pathfinder-GTMBA time-series, implying a degree of autocorrelation and some under-estimation of the trend confidence intervals. In other words, the trend in post-1995 Pathfinder SST relative to GTMBA may be of order 5 mK year^{-1} as indicated, or this magnitude may have arisen by chance from the level of low-frequency variability in the time-series.

As explained in the previous section, only the Pathfinder time series has been deseasonalised (and thus, at the same time, bias-corrected to match GTMBA in the mean). This was done because the Pathfinder-GTMBA differences were clearly regionally variable, with median difference varying

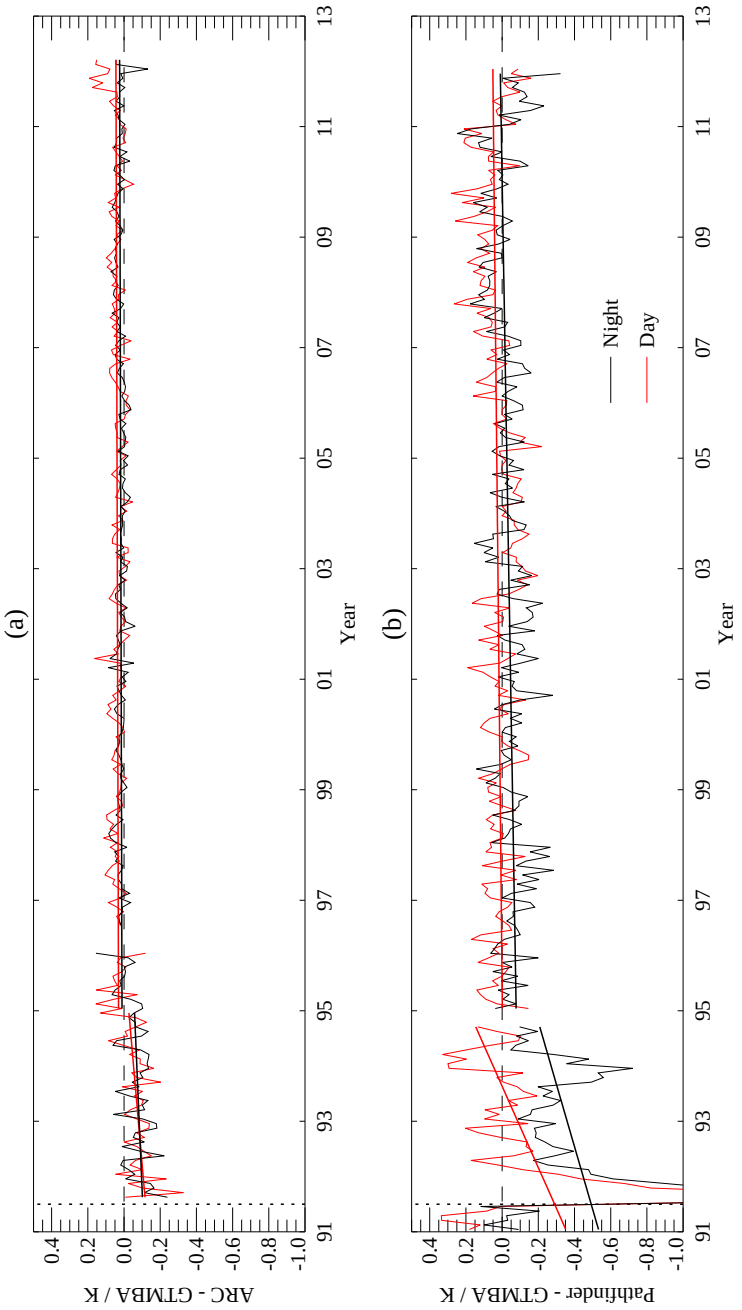

Figure 5.17: Time series of monthly median satellite - in situ differences for (a) harmonised ARC SSTs (b) deseasonalised Pathfinder v5.2. Deseasonalising removes both seasonal and regional biases relative to GTMBA meaning that only relative changes are shown in the plot. For ARC SSTs actual differences are shown.

from buoy to buoy, and were seasonally variable for some locations. The ARC-GTMBA differences were not noticeably affected by such variability and were neither deseasonalised nor bias adjusted to match GTMBA. Deseasonalising the ARC time series results in a slightly smaller trend and narrower confidence intervals: -0.9 to 0.9 mK year^{-1} during day, and -0.5 to 1.0 mK year^{-1} at night. These largely overlap with the 95% confidence intervals for the raw ARC-GTMBA trends shown in Table 5.4.

Colleagues at the National Oceanography Centre performed an independent comparison of the ARC SSTs to GTMBA data which fully accounted for autocorrelation errors and included extensive quality control of input matchups. Their work, summarised in (Merchant et al., 2012), found no significant ARC-GTMBA trends with 95% confidence intervals of -2.6 to 1.5 mK year^{-1} (daytime) and -1.8 to 1.9 mK year^{-1} (nighttime) for 1993 onwards. Their confidence intervals are wider than the ones calculated in the present work, mainly because they were working with fewer matches and also because potential autocorrelation was accounted for in their more sophisticated approach. Their having fewer matches arose through a combination of the stricter quality control and use of GTMBA measurements extracted from the International Comprehensive Ocean-Atmosphere Data set (ICOADSv2.5; Woodruff et al., 2011) rather than the high temporal resolution delayed mode data from PMEL used in this work. Despite differences in the sample sizes and methodology, both analyses agree that there is no significant trend in ARC SSTs relative to GTMBA, and that the stability of equatorial ARC SSTs after 1995 is within 3 mK year^{-1} of zero.

Prior to 1995, the ARC SSTs do not meet the stability target as seen in Table 5.5. Data for this period come from ATSR-1 and appear to be biased cold by -0.1 to -0.05 K relative to GTMBA and later sensors (Fig. 5.17) and subject to a trend between -10 and 50 mK year^{-1}. There are two reasons why the earlier ATSR-1 data may be biased relative to later data. Firstly, the ATSR-1 3.7 µm channel could not be harmonised as it was only available for the first ~8 months (see subsection 5.5.3). Secondly, the ARC SSTs could have a small residual sensitivity to stratospheric aerosol from the Mount Pinatubo eruption. However, the magnitude ~0.1 K is much smaller than for other satellite SST datasets (>1 K in Fig. 5.17b).

5.7 Summary

In this paper we describe the method used to harmonise data from the three ATSR instruments into a single climate-quality record of SST. By cross-calibrating the instrument BTs during periods of overlap (accounting for real differences in spectral response functions between sensors using radiative transfer simulations) we bring the record of BTs for all the sensors into alignment with the 3.7 and 11 µm channels of AATSR. Since, in turn, retrieval coefficients for SST also account for the different spectral responses of the instruments, this gives a harmonised record of SST.

As the ATSR instruments were designed to make accurate and stable measurements of SST they are well calibrated and characterised. This allows SST retrieval schemes which are independent of in situ measurements to be defined by RT simulation. However, as SST retrieval coefficients

generally amplify RT biases, there were discrepancies of ~0.1 K between ATSR-2 and AATSR in the unharmonised SST retrievals, implying calibration/characterisation errors of <0.1 K in the BTs. For ATSR-1 the situation was further compounded as the instrument was forced to operate at higher temperatures than intended which affected its calibration and characterisation.

When designing retrievals based on RT modelling it is useful to consider the instrument calibration and characterisation error separately. The calibration error refers to the instrument returning a correctly calibrated measurement of what it observes. The characterisation error refers to our understanding of what the instrument is observing; for the ATSRs this is the characterisation of the instrument SRFs used in the RT model. Due to the two-point calibration design of the ATSR instruments, the calibration error is likely smaller than the characterisation error. For instance, when the ATSR-1 instrument was operating with a detector temperature of 110 K (instead of the design temperature of 82 K) the calibration error was less than 15 mK, but the characterisation error resulted in a RT bias of nearly 150 mK.

During the periods of overlap between sensors we can compare the observed instrument differences against the simulated instrument differences. These double differences allow us to estimate the bias due to characterisation and forward model errors; resulting in an adjustment we can apply to the RT simulations used to generate the SST retrieval coefficients. (In the case of ATSR-1, this procedure for this is complicated by the fact of the 3.7 µm channel failure.) The new overlap-adjusted retrieval coefficients produce a harmonised SST record which is much more consistent across sensors.

Comparison against stable in situ measurements from tropical moored buoys in the Pacific (the TAO/PIRATA component of the GTMBA) shows no significant relative drift from 1995 to the end of AATSR operations in 2012. The 95% confidence interval indicates that any drift is less than 2 mK year^{-1}, surpassing the target stability of 5 mK year^{-1}. So far, this stability has only been confirmed for the tropical Pacific, due to the lack of long-term stable reference points outside this region: only the GTMBA provides the required stability in combination with records of >15 years. Data prior to 1995 come from the ATSR-1 sensor which is biased cold by -0.1 to -0.05 K (relative to later sensors and GTMBA) and is subject to a trend between -10 and 50 mK year^{-1}.

The ARC dataset provides a 20 year record of SST from space that is independent of in situ, consistent between sensors, and includes both skin SSTs and depth SSTs at a standardised local time. From 1995 it remains stable relative to tropical moorings, providing over 15 years of observations with stability apparently exceeding climate user requirements (GCOS, 2011). The ARC dataset is a significant, independent verification of changes in global SST seen in in situ datasets such as HadSST3 (Kennedy et al., 2011a,b) as shown in Merchant et al. (2012). ARC SSTs should have many applications in climate science where stability of measurement is important.

5.8 Acknowledgments

The preparation of this manuscript was funded by the UK National Centre for Earth Observation (NCEO). The ARC project was funded by the UK Natural Environment Research Council (NERC, reference NE/D001129/1) and the UK's Department of Energy and Climate Change (reference CPEG31) and Ministry of Defence. Production of ARC SSTs from January 2010 onwards was funded by the European Space Agency (ESA) Climate Change Initiative (CCI) for SST (SST-CCI). ARC SSTs are hosted and made available by the NERC Earth Observation Data Centre (http://neodc.nerc.ac.uk). In this study, we used AVHRR Pathfinder Version 5.2 (PFV5.2) data, obtained from the US National Oceanographic Data Center and GHRSST (http://pathfinder.nodc.noaa.gov). The PFV5.2 data are an updated version of the Pathfinder Version 5.0 and 5.1 collection described in Casey et al. (2010). Global Tropical Moored Buoy Array (GTMBA) data were obtained from the TAO Project Office of NOAA/PMEL.

Chapter 6

Conclusions

6.1 Introduction

In this thesis I have presented the techniques used to produce a new climate-quality SST dataset using observations from the series of ATSR instruments which were operational between 1991 and 2012. The ARC dataset is the most accurate and stable satellite-based SST product currently available, meeting both ARC and GCOS objectives.

This final chapter reviews the results presented, considers the "lessons learned" from the research, and aims to put the work into a wider context.

Among the "essential climate variables" identified by the Global Climate Observing System (GCOS, 2011) as required to give adequate monitoring of the climate surface to meet the needs of society for climate information, arguably none is more important than sea surface temperature. Accurate fields of SST are used as a boundary condition for both NWP and ocean forecasting models, which are in turn used for numerous applications including weather forecasting, fisheries information, military operations, and maritime safety. NWP re-analyses usually have been made from prescribed SST fields, and emerging coupled re-analyses are equally dependent on SST observations, since (for the present) this is thought more accurate and practical than addressing SST through direct assimilation of radiances. In climate change research, coupled atmosphere-ocean models used SST to estimate fluxes of heat and gases between the ocean and atmosphere. Furthermore measurements of SST are themselves a major component of the observational record used to assess climate change.

Data to be used for climate-change studies must be both accurate and stable. The ARC project set ambitious objectives for record accuracy and stability while requiring independence from in situ data as outlined in Merchant et al. (2008c) and chapter 1. More recently GCOS have set comparable accuracy and stability requirements for satellite-SST as an ECV (GCOS, 2011).

6.2 Radiative Transfer Basis

Radiative Transfer models can provide a sound physical basis for SST retrievals. Physically based SST retrievals can be made independent of in situ measurements, an essential objective for the ARC project. The importance of this was recognised in the design of the ATSR instrument (Harries et al., 1983) and the original SST retrieval algorithm (Závody et al., 1995). However, the initial ATSR SST retrievals were hampered by limitations in the RT models available at the time: in particular the parameterisation of water vapour continuum absorption (Clough et al., 1989). The last major update of the RT basis used for ATSR retrievals prior to this work was over a decade ago (Merchant et al., 1999). As Merchant and Le Borgne (2004) demonstrate, biases in RT simulations are generally amplified into larger SST retrieval biases when using physically based methods by a factor of order 2 to 10, depending on the retrieval context, channels used, etc. Therefore, we must ensure the RT model is as accurate as possible — ideally of order 0.01 to 0.05 K in order to make SST accurate to better than 0.1 K.

In chapter 2 I describe an updated set of RT simulations based on modern RT models, spectroscopy, and a new atmospheric profile dataset. These are shown to be in good agreement with ATSR-2 and AATSR observations in subsection 2.5.4, where plots of BT difference distributions (Fig. 2.10 and Fig. 2.11) show agreement to ~0.1 K. Figure 2.13 shows that the new RT simulations can reduce SST retrieval biases to ~0.1–0.2 K (without using the new coefficient design from chapter 3) indicating that the RT simulations are accurate to better than 0.1 K.

6.3 SST retrieval method

SST is routinely retrieved from infrared instruments using a linear (or nearly linear) function of the observed BTs where the observations are multiplied by a set of coefficients. This method is simple and effective; however it is subject to regionally varying biases due to prior and non-linearity errors (Merchant et al., 2006b) which can be of order 0.5 K. Even when combined with the updated RT simulations from chapter 2, the coefficient scheme from Závody et al. (1995) is affected by regional biases up to 0.2 K. Furthermore, few existing retrieval schemes have been characterised in terms of their sensitivity water vapour and sensitivity to SST (Merchant et al., 2009).

In chapter 3 I review the variables which resulted in biases in SST retrievals based on the Závody et al. (1995) coefficient scheme. The primary factors were atmospheric water vapour (subsection 3.3.3) and viewing geometry — the Závody scheme did not account for the fact that the relation between the nadir and forward view angles varies throughout the orbit (subsection 3.3.2). I then stratified the ARC coefficients by these variables such that the prior and non-linearity errors are minimised. I examined the simulated performance of the ARC coefficients following the criteria of (Merchant et al., 2009). The simulated regional biases (Fig. 3.8) show the new coefficient design can meet the accuracy target of better than 0.1 K. Figure 3.13 shows that the coefficients have minimal sensitivity to TCWV, while Fig. 3.14 shows the sensitivity to true SST is within 1% of ideal.

With my improved coefficient design the residual forward model errors became apparent as a ~0.1 K different between the two channel and three channel retrievals (section 3.4). In order to resolve these discrepancies I used a mutual bias correction to a reference retrieval, ensuring inter-algorithm consistency without reference to any in situ SST. Ordinarily the D3 retrieval would be the obvious choice as a reference retrieval, but due to the uncertainty in the calibration of the AATSR 12 µm channel (see O'Hara et al., 2013) I created an alternative retrieval using dual-view 3.7 and 11 µm channels referred to as D2*. By tying the ARC SST retrievals to D2* I eliminated the remaining discrepancies due to residual forward model errors and ensured that the ARC SST retrievals are not affected by the uncertainty in the AATSR 12 µm channel.

6.4 Initial validation

In chapter 4 I present a comparison of the initial (unharmonised) ARC SST products against in situ drifters and tropical moored buoys. Frequently the validation of satellite SST products ignores the fact that satellites are not sensitive to the same SST as in situ measurements (section 1.4) as the differences are small compared to the accuracy of typical satellite products. However, when aiming for accuracy of order 0.1 K the skin effect and diurnal cycle cannot be neglected. Section 4.3 examines the skin effect and diurnal cycle; identifying both factors in the satellite-drifter differences and validating the performance of the Met Office skin and diurnal model used to model them.

Section 4.4 presents a comparison of skin and depth adjusted ARC SSTs against global in situ measurements. I show that the regional bias target is met for AATSR SSTs and present evidence that the ARC SSTs are more precise than in situ drifter measurements.

6.5 Harmonisation

In chapter 5 I develop a inter-satellite BT harmonisation technique to give consistency and stability across multiple sensors. By cross-calibrating the BTs during periods of overlap I brought the earlier sensors into alignment with AATSR. Comparison against stable in situ measurements from tropical moored buoys shows agreement between the two records after 1995 was within 2 mK year^{-1} to 95% confidence meeting the stability target.

As the harmonisation was performed in BT space it is not restricted to SST retrieval. Other retrievals which use clear-sky BTs over a similar temperature range would benefit from the BT harmonisation. The BT harmonisation has also contributed to the analysis of the AATSR 12 µm anomaly (O'Hara et al., 2013).

6.6 Independent Validation

Due to the stringent targets set for the ARC SST products and their intended use as a reference dataset it is important that users can have confidence in the ARC SSTs. Therefore the ARC SSTs were "independently" validated by teams working at the University of Leicester, the UK Met Office, and the National Oceanography Centre. Their assessment is documented in Merchant et al. (2012) and confirms the accuracy, precision, and stability of the ARC SSTs.

6.7 Influence on Other Work

6.7.1 ARC_L2P

Operational AATSR SSTs were used as a reference field to bias correct the Met Office Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system (Donlon et al., 2012b); OSTIA SSTs are in turn used as the boundary condition in weather forecasts at the Met Office and ECMWF. When the advantages of the ARC SST products became apparent through my results presented here and independent assessment at the Met Office there was a demand for the operational AATSR products to be upgraded to match the ARC products.

This was done in two ways. Firstly, I modified my software for operational use within the ESA ground segment which handles AATSR data. This includes near real time (NRT) capability and producing output files compliant with GHRSST GDS 2.0 (GHRSST Science Team, 2012). Secondly, in order to support legacy users, I produced coefficient files compatible with the old processing software (i.e. compatible with the Závody et al., 1995, scheme) using my updated RT simulations (chapter 2).

These updates were scheduled to be implemented in ground segment in summer 2012. Unfortunately the Envisat satellite, which carries the AATSR instrument, suddenly ceased transmission on 8th April 2012 and was declared lost a month later. However, the updates were made in February 2013 in order to support the next ESA reprocessing of the (A)ATSR Multi-mission Archive.

6.7.2 SLSTR

The future Sea and Land Surface Temperature Radiometer (SLSTR) is a dual-view radiometer building on the heritage of the ATSR instruments and part of the payload of Sentinel-3 (Donlon et al., 2012a) due to be launched in 2014. The SST retrieval scheme will be based on the one I designed for the ARC project (chapter 2 and chapter 3). Pre-launch SST retrieval coefficients for the SLSTR instrument were generated by S. MacCallum using the software I developed for the ARC project.

6.7.3 ESA CCI

The European Space Agency Climate Change Initiative (ESA CCI) for SST (SST-CCI) project aims to produce global satellite SST products meeting GCOS (2011) ECV requirements using multi-sensor

data archives. In the first phase the primary sensors are the ATSRs and AVHRRs: the AVHRRs providing high spatial coverage; and the ATSRs providing an accurate and stable, but sparse, baseline. In order to produce a climate data record (CDR) independent of in situ SST, the SST-CCI project used ARC SSTs as the reference. Due to the importance of the ARC SST products, ESA funded the production of ARC SSTs from January 2010 to the end of the AATSR mission under SST-CCI.

The ARC processing software was reused within SST-CCI; adding support for AVHRR sensors, and optimal estimation techniques.

6.7.4 Other

- ARC-Lake is an ESA funded project to measure Lake Surface Water Temperature (LWST) and Lake Ice Cover from the ATSR instruments. The ARC-Lake project reused the infrastructure I built for processing the ATSR data archive at the NERC Earth Observation Data Centre. The ARC processing software was modified to support cloud clearing over lakes, and retrieval of LWST (as opposed to SST).

- ARC SST data are being used to assess historical in situ datasets such as HadSST3, providing independent corroboration of the trends observed in recent decades (Merchant et al., 2012)

- The ARC SST products have been made available in CMIP5 format for climate modellers via obs4MIPS. The data are available from the STFC Centre for Environmental Data Archival (CEDA) Earth System Grid Federation (ESGF).

- ARC SSTs will be used at the National Oceanography Centre to assess biases and uncertainties in modern in situ SST measurements in an attempt to better understand the biases and uncertainties present in the historical record.

6.8 Future Work

6.8.1 ATSR-1

SST retrievals from the ATSR-1 instrument are of lower quality than those from the later two instruments. This was due to three main factors: the ATSR-1 instrument was operating at much higher temperatures than intended, the 3.7 μm channel failed early during the instruments operation, and the effects of the Mount Pinatubo eruption in 1991. This work attempted to address these issues, but the resulting ARC SSTs prior to 1995 did not meet the stability target. Work to improve the ATSR-1 SST retrieval using independent measurements of aerosol optical depths and further investigation of the instrument stability has been proposed for ESA SST-CCI Phase II

6.8.2 Instrument calibration

While the calibration and characterisation of the ATSR instruments are good, they are not perfect. In particular, the AATSR 12 μm channel appears to be affected by a ~0.2 K bias which has been investigated by the AATSR Anomaly Review Board (ARB; O'Hara et al., 2013). The investigation concluded that a shift of the 12 μm spectral response function along with an error in the non-linearity correction could explain the observed discrepancy between ATSR-2 and AATSR (section 5.4). Accounting for this upfront should lead to improved SST retrievals.

6.8.3 Extension to SLSTR

The ARC SST record extends to April 2012 when the Envisat satellite, which carried the AATSR instrument, failed and there will be no comparable dual-view radiometer operating in space until the launch of Sentinel-3, carrying the SLSTR instrument, which is due in 2015. The SST retrieval scheme used for SLSTR will be based on the work presented in this thesis; however, the gap of more than three years between AATSR and SLSTR means the SST record cannot be harmonised using coincident observations between the two instruments. Work to bridge the gap between the instruments using the AVHRR and Infrared Atmospheric Sounding Interferometer (IASI) carried onboard the MetOp satellites has been proposed for ESA SST-CCI Phase II. The AVHRRs are operational meteorological instruments which measure SST; however, they are not as well calibrated as the ATSRs so the IASI instrument will be used to monitor the stability and calibration of AVHRR during the gap.

6.8.4 Ongoing Reprocessing

Since the completion of the ARC project several of the input datasets and models used to generate the ARC SSTs have been updated. The input satellite data were the BT observations from with ESA's (A)ATSR Multi-mission Archive. This archive has recently been reprocessed adding improved geolocation, nadir-forward view collocation, and reflectance channel calibration. The ERA-40 data used as NWP for the ARC processing have since been superseded by ERA-interim. The fast forward model, RTTOV, used within the ARC processing has been updated since the dataset was generated. Work to verify the benefit of these updates been proposed for ESA SST-CCI Phase II.

6.9 Summary

This project has produced a new >20 year dataset of SST from ATSR suitable for climate research. It is the first satellite-based record of SST which is independent of in situ, consistent between different sensors, and consistent between different channel combinations. The ARC data are very precise (low noise) estimated at 0.14 K from three-way precision analysis for 2003 through 2009 which is lower than the global drifter network. The ARC data are highly accurate, with regional biases relative to drifting buoys lower than 0.1 K. The ARC data are very stable with drift relative to tropical moorings

less than 2 mK year^{-1} to 95% confidence interval. The ARC SST data meet the ARC project objectives and GCOS objectives and are expected to have many applications in climate science.

The techniques and software developed in this project have proved successful and have been used in several subsequent research projects.

Bibliography

Allen, M., C. Mutlow, G. Blumberg, J. Christy, R. McNider, and D. Llewellyn-Jones (1994). Global change detection. *Nature*, **370** (6484), pages 24–24. doi: 10.1038/370024b0.

Anding, D., and R. Kauth (1970). Estimation of sea surface temperature from space. *Remote Sensing of Environment*, **1** (4), pages 217–220. doi: 10.1016/S0034-4257(70)80002-5.

Baran, A. J., and J. S. Foot (1994). New application of the operational sounder HIRS in determining a climatology of sulphuric acid aerosol from the pinatubo eruption. *Journal of Geophysical Research*, **99** (D12), pages 25673–25679. doi: 10.1029/94JD02044.

Barton, I. J. (1995). Satellite-derived sea surface temperatures: Current status. *Journal of Geophysical Research*, **100** (C5), pages 8777–8790. doi: 10.1029/95JC00365.

Barton, I. J. (1998). Improved techniques for the derivation of sea surface temperatures from ATSR data. *Journal of Geophysical Research*, **103** (C4), pages 8139–8152. doi: 10.1029/97JC02569.

Bertie, J. E., and Z. Lan (1996). Infrared intensities of liquids XX: the intensity of the OH stretching band of liquid water revisited, and the best current values of the optical constants of H_2O(l) at 25°C between 15,000 and 1 cm^{-1}. *Applied Spectroscopy*, **50** (8), pages 1047–1057. URL http://as.osa.org/abstract.cfm?URI=as-50-8-1047.

Birks, A. R. (2006). Latitude dependent bias correction. AATSR technical note, Rutherford Appleton Laboratory. URL http://envisat.esa.int/pub/ESA_DOC/ENVISAT/AATSR/Latitude_Dependent_Bias_Correction.pdf.

Birks, A. R. (2008). Detector temperature dependence of the 12 micron channel filter profile and its impact on the ATSR-1 calibration. Technical note, Rutherford Appleton Laboratory.

Blackmore, T., A. O'Carroll, K. Fennig, and R. Saunders (2012). Correction of AVHRR pathfinder SST data for volcanic aerosol effects using ATSR SSTs and TOMS aerosol optical depth. *Remote Sensing of Environment*, **116**, pages 107–117. doi: 10.1016/j.rse.2011.04.040.

Blake, D. (2005). Methane, nonmethane hydrocarbons, alkyl nitrates, and chlorinated carbon compounds including 3 chlorofluorocarbons (CFC-11, CFC-12, and CFC-113) in whole-air samples. In

Trends: A Compendium of Data on Global Change. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.

Bourlès, B., R. Lumpkin, M. J. McPhaden, F. Hernandez, P. Nobre, E. Campos, L. Yu, S. Planton, A. Busalacchi, A. D. Moura, J. Servain, and J. Trotte (2008). The pirata program: History, accomplishments, and future directions. *Bulletin of the American Meteorological Society*, **89** (8), pages 1111–1125. doi: 10.1175/2008BAMS2462.1.

Brown, S. J., A. R. Harris, I. M. Mason, and A. M. Závody (1997). New aerosol robust sea surface temperature algorithms for the along-track scanning radiometer. *Journal of Geophysical Research*, **102** (C13), pages 27973–27989. doi: 10.1029/97JC01192.

Casey, K. S., T. B. Brandon, P. Cornillon, and R. Evans (2010). The past, present, and future of the AVHRR pathfinder SST program. In Barale, V., J. Gower, and L. Alberotanza, editors, *Oceanography from Space*, pages 273–287. Springer Netherlands, Dordrecht. ISBN 978-90-481-8681-5. doi: 10.1007/978-90-481-8681-5_16.

Chevallier, F. (2002). Sampled database of 60-level atmospheric profiles from the ECMWF analyses. Technical Report 4, ECMWF, Reading, UK. URL `http://www.ecmwf.int/publications/library/do/references/show?id=83287`.

Clough, S. A., F. X. Kneizys, and R. W. Davies (1989). Line shape and the water vapor continuum. *Atmospheric Research*, **23** (3-4), pages 229–241. doi: 10.1016/0169-8095(89)90020-3.

Clough, S. A., M. W. Shephard, E. J. Mlawer, J. S. Delamere, M. J. Iacono, K. Cady-Pereira, S. Boukabara, and P. D. Brown (2005). Atmospheric radiative transfer modeling: a summary of the AER codes. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **91** (2), pages 233–244. doi: 10.1016/j.jqsrt.2004.05.058.

Corlett, G. K., I. J. Barton, C. Donlon, M. Edwards, S. Good, L. A. Horrocks, D. Llewellyn-Jones, C. J. Merchant, P. J. Minnett, and T. J. Nightingale (2006). The accuracy of SST retrievals from AATSR: an initial assessment through geophysical validation against in situ radiometers, buoys and other SST data sets. *Advances in Space Research*, **37** (4), pages 764–769. doi: 10.1016/j.asr.2005.09.037.

Cox, C., and W. Munk (1954). Measurement of the roughness of the sea surface from photographs of the sun's glitter. *Journal of the Optical Society of America*, **44** (11), pages 838–850. doi: 10.1364/JOSA.44.000838.

de Boyer Montégut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, **109** (C12). doi: 10.1029/2004JC002378.

Deschamps, P. Y., and T. Phulpin (1980). Atmospheric correction of infrared measurements of sea surface temperature using channels at 3.7, 11 and 12 μm. *Boundary-Layer Meteorology*, **18** (2), pages 131–143. doi: 10.1007/BF00121320.

Deshler, T., M. E. Hervig, D. J. Hofmann, J. M. Rosen, and J. B. Liley (2003). Thirty years of in situ stratospheric aerosol size distribution measurements from laramie, wyoming (41°N), using balloon-borne instruments. *Journal of Geophysical Research*, **108** (D5). doi: 10.1029/2002JD002514.

Donlon, C., B. Berruti, A. Buongiorno, M.-H. Ferreira, P. Féménias, J. Frerick, P. Goryl, U. Klein, H. Laur, C. Mavrocordatos, J. Nieke, H. Rebhan, B. Seitz, J. Stroede, and R. Sciarra (2012a). The global monitoring for environment and security (GMES) sentinel-3 mission. *Remote Sensing of Environment*, **120**, pages 37–57. doi: 10.1016/j.rse.2011.07.024.

Donlon, C., I. S. Robinson, W. Wimmer, G. Fisher, M. Reynolds, R. Edwards, and T. J. Nightingale (2008). An infrared sea surface temperature autonomous radiometer (ISAR) for deployment aboard volunteer observing ships (VOS). *Journal of Atmospheric and Oceanic Technology*, **25** (1), pages 93–113. doi: 10.1175/2007JTECHO505.1.

Donlon, C. J., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer (2012b). The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sensing of Environment*, **116**, pages 140–158. doi: 10.1016/j.rse.2010.10.017.

Donlon, C. J., P. J. Minnett, C. L. Gentemann, T. J. Nightingale, I. J. Barton, B. Ward, and M. J. Murray (2002). Toward improved validation of satellite sea surface skin temperature measurements for climate research. *Journal of Climate*, **15** (4), pages 353–369. doi: 10.1175/1520-0442(2002)015<0353:TIVOSS>2.0.CO;2.

Donlon, C. J., N. Rayner, I. Robinson, D. J. S. Poulter, K. S. Casey, J. Vazquez-Cuervo, E. Armstrong, A. Bingham, O. Arino, C. L. Gentemann, D. May, P. Le Borgne, J. Piollé, I. J. Barton, H. Beggs, C. J. Merchant, S. Heinz, A. R. Harris, G. Wick, B. Emery, P. J. Minnett, R. Evans, D. Llewellyn-Jones, C. Mutlow, R. W. Reynolds, and H. Kawamura (2007). The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *Bulletin of the American Meteorological Society*, **88** (8), pages 1197–1213. doi: 10.1175/BAMS-88-8-1197.

Downing, H. D., and D. Williams (1975). Optical constants of water in the infrared. *Journal of Geophysical Research*, **80** (12), pages 1656–1661. doi: 10.1029/JC080i012p01656.

Dudhia, A. (2006). Reference forward model (RFM). http://www.atm.ox.ac.uk/RFM/. URL http://www.atm.ox.ac.uk/RFM/.

Edwards, D. P. (1992). GENLN2: a general line-by-line atmospheric transmittance and radiance model. NCAR Technical Note NCAR/TN-367+STR, National Center for Atmospheric Research, Boulder, Colorado. URL http://nldr.library.ucar.edu/repository/technotes/TECH-NOTE-000-000-000-178.

Edwards, T., R. Browning, J. Delderfield, D. J. Lee, and K. A. Lidiard (1990). The along track scanning radiometer - measurement of sea-surface temperature from ERS-1. *Journal of the British Interplanetary Society*, **43**, pages 160–180. URL http://adsabs.harvard.edu/abs/1990JBIS...43..160E.

Embury, O., C. J. Merchant, and M. J. Filipiak (2012a). A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Basis in radiative transfer. *Remote Sensing of Environment*, **116**, pages 32–46. doi: 10.1016/j.rse.2010.10.016.

Embury, O., and C. J. Merchant (2012). A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: A new retrieval scheme. *Remote Sensing of Environment*, **116**, pages 47–61. doi: 10.1016/j.rse.2010.11.020.

Embury, O., C. J. Merchant, and G. K. Corlett (2012b). A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Initial validation, accounting for skin and diurnal variability effects. *Remote Sensing of Environment*, **116**, pages 62–78. doi: 10.1016/j.rse.2011.02.028.

Emery, W. J., D. J. Baldwin, P. Schlüssel, and R. W. Reynolds (2001). Accuracy of in situ sea surface temperatures used to calibrate infrared satellite measurements. *Journal of Geophysical Research*, **106** (C2), pages 2387–2405. doi: 10.1029/2000JC000246.

Emery, W. J., Y. Yu, G. A. Wick, P. Schluessel, and R. W. Reynolds (1994). Correcting infrared satellite estimates of sea surface temperature for atmospheric water vapor attenuation. *Journal of Geophysical Research*, **99** (C3), pages 5219–5236. doi: 10.1029/93JC03215.

Fairall, C. W., E. F. Bradley, J. S. Godfrey, G. A. Wick, J. B. Edson, and G. S. Young (1996). Cool-skin and warm-layer effects on sea surface temperature. *Journal of Geophysical Research*, **101** (C1), pages 1295–1308. doi: 10.1029/95JC03190.

Fairall, C. W., E. F. Bradley, J. E. Hare, A. A. Grachev, and J. B. Edson (2003). Bulk parameterization of Air–Sea fluxes: Updates and verification for the COARE algorithm. *Journal of Climate*, **16** (4), pages 571–591. doi: 10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2.

Friedman, D. (1969). Infrared characteristics of ocean water (1.5–15 μ). *Applied Optics*, **8** (10), pages 2073–2078. doi: 10.1364/AO.8.002073.

Gatebe, C. K., M. D. King, A. I. Lyapustin, G. T. Arnold, and J. Redemann (2005). Airborne spectral measurements of ocean directional reflectance. *Journal of the Atmospheric Sciences*, **62** (4), pages 1072–1092. doi: 10.1175/JAS3386.1.

GCOS (2011). Systematic observation requirements for satellite-based products for climate. Supplemental details to the satellite-based component of the "Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update)" GCOS-154, World Meteorological Organization. URL http://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf.

Gentemann, C. L., P. J. Minnett, P. Le Borgne, and C. J. Merchant (2008). Multi-satellite measurements of large diurnal warming events. *Geophysical Research Letters*, **35** (22). doi: 10.1029/2008GL035730.

Gentemann, C. L., P. J. Minnett, and B. Ward (2009). Profiles of ocean surface heating (POSH): a new model of upper ocean diurnal warming. *Journal of Geophysical Research*, **114** (C7). doi: 10.1029/2008JC004825.

GHRSST Science Team (2012). The recommended GHRSST data specification (GDS) 2.0. Technical Report 5, GHRSST International Project Office. URL http://www.ghrsst.org.

Good, E., X. Kong, O. Embury, C. Merchant, and J. Remedios (2012). An infrared desert dust index for the along-track scanning radiometers. *Remote Sensing of Environment*, **116**, pages 159–176. doi: 10.1016/j.rse.2010.06.016.

Grooß, J., and J. M. Russell (2005). Technical note: A stratospheric climatology for o_3, H_2O, CH_4, NOx, HCl and HF derived from HALOE measurements. *Atmospheric Chemistry and Physics*, **5** (10), pages 2797–2807. doi: 10.5194/acp-5-2797-2005.

Harries, J. E., D. T. Llewellyn-Jones, P. J. Minnett, R. W. Saunders, A. M. Zavody, P. Wadhams, P. K. Taylor, and J. T. Houghton (1983). Observations of sea-surface temperature for climate research [and discussion]. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **309** (1508), pages 381–395. doi: 10.2307/37364. URL http://www.jstor.org/stable/37364.

Harris, A. R., M. A. Saunders, J. S. Foot, K. F. Smith, and C. T. Mutlow (1995). Improved sea surface temperature measurements from space. *Geophysical Research Letters*, **22** (16), pages 2159–2162. doi: 10.1029/95GL02019.

Hegerl, G. C., and N. L. Bindoff (2005). OCEAN SCIENCE: warming the world's oceans. *Science*, **309** (5732), pages 254–255. doi: 10.1126/science.1114456.

Hess, M., P. Koepke, and I. Schult (1998). Optical properties of aerosols and clouds: The software package OPAC. *Bulletin of the American Meteorological Society*, **79** (5), pages 831–844. doi: 10.1175/1520-0477(1998)079<0831:OPOAAC>2.0.CO;2.

Horrocks, L. A., B. Candy, T. J. Nightingale, R. W. Saunders, A. G. O'Carroll, and A. R. Harris (2003a). Parameterizations of the ocean skin effect and implications for satellite-based measurement of sea-surface temperature. *Journal of Geophysical Research*, **108** (C3). doi: 10.1029/2002JC001503.

Horrocks, L. A., A. R. Harris, and R. W. Saunders (2003b). Modelling the diurnal thermocline for daytime bulk SST from AATSR. Forecasting Research Technical Report 418, Met Office. URL http://research.metoffice.gov.uk/research/nwp/publications/papers/technical_reports/2003/FRTR418/FRTR418.pdf.

Huber, P. J. (1981). *Robust Statistics*. Wiley, New York. ISBN 9780471418054.

Hwang, P. A. (1997). A study of the wavenumber spectra of short water waves in the ocean. part II: spectral model and mean square slope. *Journal of Atmospheric and Oceanic Technology*, **14** (5), pages 1174–1186. doi: 10.1175/1520-0426(1997)014<1174:ASOTWS>2.0.CO;2.

Ignatov, A., I. Laszlo, E. D. Harrod, K. B. Kidwell, and G. P. Goodrum (2004). Equator crossing times for NOAA, ERS and EOS sun-synchronous satellites. *International Journal of Remote Sensing*, **25** (23), pages 5255–5266. doi: 10.1080/01431160410001712981.

Immler, F. J., J. Dykema, T. Gardiner, D. N. Whiteman, P. W. Thorne, and H. Vömel (2010). Reference quality upper-air measurements: guidance for developing GRUAN data products. *Atmospheric Measurement Techniques*, **3** (5), pages 1217–1231. doi: 10.5194/amt-3-1217-2010.

Jin, M., and R. E. Treadon (2003). Correcting the orbit drift effect on AVHRR land surface skin temperature measurements. *International Journal of Remote Sensing*, **24** (22), pages 4543–4558. doi: 10.1080/0143116031000095943.

Kantha, L. H., and C. A. Clayson (1994). An improved mixed layer model for geophysical applications. *Journal of Geophysical Research*, **99** (C12), pages 25235–25266. doi: 10.1029/94JC02257.

Kaplan, A., M. A. Cane, Y. Kushnir, A. C. Clement, M. B. Blumenthal, and B. Rajagopalan (1998). Analyses of global sea surface temperature 1856–1991. *Journal of Geophysical Research*, **103** (C9), page 18567. doi: 10.1029/97JC01736.

Kearns, E. J., J. A. Hanafin, R. H. Evans, P. J. Minnett, and O. B. Brown (2000). An independent assessment of pathfinder AVHRR sea surface temperature accuracy using the marine atmosphere emitted radiance interferometer (MAERI). *Bulletin of the American Meteorological Society*, **81** (7), pages 1525–1536. doi: 10.1175/1520-0477(2000)081<1525:AIAOPA>2.3.CO;2.

Keeling, C. D., and T. P. Whorf (2005). Atmospheric CO_2 records from sites in the SIO air sampling network. In *Trends: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.

Kennedy, J., R. Smith, and N. Rayner (2012). Using AATSR data to assess the quality of in situ sea-surface temperature observations for climate studies. *Remote Sensing of Environment*, **116**, pages 79–92. doi: 10.1016/j.rse.2010.11.021.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby (2011a). Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. measurement and sampling uncertainties. *Journal of Geophysical Research*, **116** (D14). doi: 10.1029/2010JD015218.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby (2011b). Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. biases and homogenization. *Journal of Geophysical Research*, **116** (D14). doi: 10.1029/2010JD015220.

Kent, E. C., J. J. Kennedy, D. I. Berry, and R. O. Smith (2010). Effects of instrumentation changes on sea surface temperature measured in situ. *Wiley Interdisciplinary Reviews: Climate Change*, **1** (5), pages 718–728. doi: 10.1002/wcc.55.

Kiehl, J. T., and K. E. Trenberth (1997). Earth's annual global mean energy budget. *Bulletin of the American Meteorological Society*, **78** (2), pages 197–208. doi: 10.1175/1520-0477(1997)078<0197:EAGMEB>2.0.CO;2.

Kilpatrick, K. A., G. P. Podestá, and R. Evans (2001). Overview of the NOAA/NASA advanced very high resolution radiometer pathfinder algorithm for sea surface temperature and associated matchup database. *Journal of Geophysical Research*, **106** (C5), pages 9179–9197. doi: 10.1029/1999JC000065.

Koepke, P., M. Hess, I. Schult, and E. P. Shettle (1997). Global aerosol data set. Technical Report 243, Max-Planck-Institut für Meteorologie, Hamburg.

Lambert, A., R. G. Grainger, C. D. Rodgers, F. W. Taylor, J. L. Mergenthaler, J. B. Kumer, and S. T. Massie (1997). Global evolution of the mt. pinatubo volcanic aerosols observed by the infrared limb-sounding instruments CLAES and ISAMS on the upper atmosphere research satellite. *Journal of Geophysical Research*, **102** (D1), page 1495. doi: 10.1029/96JD00096.

Le Borgne, P., A. Marsouin, F. Orain, and H. Roquet (2012). Operational sea surface temperature bias adjustment using AATSR data. *Remote Sensing of Environment*, **116**, pages 93–106. doi: 10.1016/j.rse.2010.02.023.

Le Treut, H., R. Somerville, U. Cubasch, Y. Ding, C. Mauritzen, A. Mokssit, T. Peterson, and M. Prather (2007). Historical overview of climate change. In Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, editors, *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 93–128. Cambridge University Press, Cambridge, United Kingdom.

Li, X., W. Pichel, E. Maturi, P. Clemente-Colón, and J. Sapper (2001). Deriving the operational nonlinear multichannel sea surface temperature algorithm coefficients for NOAA-15 AVHRR/3. *International Journal of Remote Sensing*, **22** (4), pages 699–704. doi: 10.1080/01431160010013793.

Liang, X., and A. Ignatov (2011). Monitoring of IR clear-sky radiances over oceans for SST (MICROS). *Journal of Atmospheric and Oceanic Technology*, **28** (10), pages 1228–1242. doi: 10.1175/JTECH-D-10-05023.1.

Liu, Y., M. Su, X. Yan, and W. T. Liu (2000). The Mean-Square slope of ocean surface waves and its effects on radar backscatter. *Journal of Atmospheric and Oceanic Technology*, **17** (8), pages 1092–1105. doi: 10.1175/1520-0426(2000)017<1092:TMSSOO>2.0.CO;2.

Lukas, R. (1991). The diurnal cycle of sea surface temperature in the western equatorial pacific. In *TOGA Notes*, volume 2, pages 1–5.

Lumpkin, R., and M. Pazos (2007). Measuring surface currents with surface velocity program drifters: the instrument, its data, and some recent results. In Griffa, A., A. D. Kirwan, A. J. Mariano, T. Özgökmen, and H. T. Rossby, editors, *Lagrangian analysis and prediction of coastal and ocean dynamics*. Cambridge University Press, Cambridge. ISBN 9780521870184.

Mackie, S., C. J. Merchant, O. Embury, and P. Francis (2010). Generalized bayesian cloud detection for satellite imagery. part 2: Technique and validation for daytime imagery. *International Journal of Remote Sensing*, **31** (10), pages 2595–2621. doi: 10.1080/01431160903051711.

Marmorino, G. O., and G. B. Smith (2005). Bright and dark ocean whitecaps observed in the infrared. *Geophysical Research Letters*, **32** (11). doi: 10.1029/2005GL023176.

Mason, G. (1991). *ATSR Test and Calibration Report. Satellite Measurement Of Sea Surface Temperature*. D.Phil, Oxford University.

Mason, I. M., P. H. Sheather, J. A. Bowles, and G. Davies (1996). Blackbody calibration sources of high accuracy for a spaceborne infrared instrument: the along track scanning radiometer. *Applied Optics*, **35** (4), page 629. doi: 10.1364/AO.35.000629.

Massie, S. T., T. Deshler, G. E. Thomas, J. L. Mergenthaler, and J. M. Russell (1996). Evolution of the infrared properties of the mount pinatubo aerosol cloud over laramie, wyoming. *Journal of Geophysical Research*, **101** (D17), pages 23007–23019. doi: 10.1029/96JD01991.

Masuda, K. (2006). Infrared sea surface emissivity including multiple reflection effect for isotropic gaussian slope distribution model. *Remote Sensing of Environment*, **103** (4), pages 488–496. doi: 10.1016/j.rse.2006.04.011.

Masuda, K., T. Takashima, and Y. Takayama (1988). Emissivity of pure and sea waters for the model sea surface in the infrared window regions. *Remote Sensing of Environment*, **24** (2), pages 313–329. doi: 10.1016/0034-4257(88)90032-6.

Matricardi, M. (2007). An inter-comparison of line-by-line radiative transfer models. Technical Report 525, ECMWF, Reading. URL `http://www.ecmwf.int/publications/library/do/references/show?id=87777`.

Matthiesen, S., and C. J. Merchant (2003). Sea surface temperatures in marginal ice zones. Final report on pathfinder project, University of Edinburgh, Edinburgh, UK.

Maury, M. F. (1858). *Explanations and Sailing Directions to Accompany the Wind and Current Charts*, volume 1. Cornelius Wendell.

Maury, M. F. (1859). *Explanations and Sailing Directions to Accompany the Wind and Current Charts*, volume 2. Cornelius Wendell.

McPhaden, M., K. Ando, B. Bourlès, H. Freitag, R. Lumpkin, Y. Masumoto, V. Murty, P. Nobre, M. Ravichandran, J. Vialard, D. Vousden, and W. Yu (2010). The global tropical moored buoy array. In *Sustained Ocean Observations and Information for Society*, volume 2, pages 668–682. European Space Agency, Venice, Italy. ISBN 978-3-86987-200-1. doi: 10.5270/OceanObs09.cwp.61.

McPhaden, M. J. (1995). The tropical atmosphere ocean (TAO) array is completed. *Bulletin of the American Meteorological Society*, **76** (5), pages 739–741.

McPhaden, M. J., G. Meyers, K. Ando, Y. Masumoto, V. S. N. Murty, M. Ravichandran, F. Syamsudin, J. Vialard, L. Yu, and W. Yu (2009). RAMA: the research moored array for African–Asian–Australian monsoon analysis and prediction. *Bulletin of the American Meteorological Society*, **90** (4), pages 459–480. doi: 10.1175/2008BAMS2608.1.

Merchant, C. J., O. Embury, P. Le Borgne, and B. Bellec (2006a). Saharan dust in nighttime thermal imagery: Detection and reduction of related biases in retrieved sea surface temperature. *Remote Sensing of Environment*, **104** (1), pages 15–30. doi: 10.1016/j.rse.2006.03.007.

Merchant, C. J., O. Embury, N. A. Rayner, D. I. Berry, G. K. Corlett, K. Lean, K. L. Veal, E. C. Kent, D. T. Llewellyn-Jones, J. J. Remedios, and R. Saunders (2012). A 20 year independent record of sea surface temperature for climate from along-track scanning radiometers. *Journal of Geophysical Research*, **117** (C12). doi: 10.1029/2012JC008400.

Merchant, C. J., M. J. Filipiak, P. Le Borgne, H. Roquet, E. Autret, J. Piollé, and S. Lavender (2008a). Diurnal warm-layer events in the western mediterranean and european shelf seas. *Geophysical Research Letters*, **35** (4). doi: 10.1029/2007GL033071.

Merchant, C. J., and A. R. Harris (1999). Toward the elimination of bias in satellite retrievals of sea surface temperature 2. comparison with in situ measurements. *Journal of Geophysical Research*, **104** (C10), pages 23579–23590. doi: 10.1029/1999JC900106.

Merchant, C. J., A. R. Harris, E. Maturi, and S. N. MacCallum (2005). Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval. *Quarterly Journal of the Royal Meteorological Society*, **131** (611), pages 2735–2755. doi: 10.1256/qj.05.15.

Merchant, C. J., A. R. Harris, M. J. Murray, and A. M. Závody (1999). Toward the elimination of bias in satellite retrievals of sea surface temperature 1. theory, modeling and interalgorithm comparison. *Journal of Geophysical Research*, **104** (C10), pages 23565–23578. doi: 10.1029/1999JC900105.

Merchant, C. J., A. R. Harris, H. Roquet, and P. Le Borgne (2009). Retrieval characteristics of non-linear sea surface temperature from the advanced very high resolution radiometer. *Geophysical Research Letters*, **36** (17). doi: 10.1029/2009GL039843.

Merchant, C. J., L. A. Horrocks, J. R. Eyre, and A. G. O'Carroll (2006b). Retrievals of sea surface temperature from infrared imagery: origin and form of systematic errors. *Quarterly Journal of the Royal Meteorological Society*, **132** (617), pages 1205–1223. doi: 10.1256/qj.05.143.

Merchant, C. J., and P. Le Borgne (2001). Designing OSI-SAF surface temperature retrievals to be robust to stratospheric aerosol. In *Proceedings The 2001 EUMETSAT Meteorological Satellite Data Users' Conference*. Eumetsat, Antalya, Turkey. ISBN 9789291100446.

Merchant, C. J., and P. Le Borgne (2004). Retrieval of sea surface temperature from space, based on modeling of infrared radiative transfer: Capabilities and limitations. *Journal of Atmospheric and Oceanic Technology*, **21** (11), pages 1734–1746. doi: 10.1175/JTECH1667.1.

Merchant, C. J., P. Le Borgne, A. Marsouin, and H. Roquet (2008b). Optimal estimation of sea surface temperature from split-window observations. *Remote Sensing of Environment*, **112** (5), pages 2469–2484. doi: 10.1016/j.rse.2007.11.011.

Merchant, C. J., D. Llewellyn-Jones, R. W. Saunders, N. Rayner, E. C. Kent, C. P. Old, D. I. Berry, A. Birks, T. Blackmore, G. K. Corlett, O. Embury, V. L. Jay, J. Kennedy, C. T. Mutlow, T. J. Nightingale, A. G. O'Carroll, M. J. Pritchard, J. J. Remedios, and S. Tett (2008c). Deriving a sea surface temperature record suitable for climate change research from the along-track scanning radiometers. *Advances in Space Research*, **41** (1), pages 1–11. doi: 10.1016/j.asr.2007.07.041.

Merchant, C. J., J. J. Simpson, and A. R. Harris (2003). A cross-calibration of GMS-5 thermal channels against ATSR-2. *Remote Sensing of Environment*, **84** (2), pages 268–282. doi: 10.1016/S0034-4257(02)00112-8.

Minnett, P. J. (1986). A numerical study of the effects of anomalous north atlantic atmospheric conditions on the infrared measurement of sea surface temperature from space. *Journal of Geophysical Research*, **91** (C7), pages 8509–8521. doi: 10.1029/JC091iC07p08509.

Minnett, P. J. (1991). Consequences of sea surface temperature variability on the validation and applications of satellite measurements. *Journal of Geophysical Research*, **96** (C10), pages 18475–18489. doi: 10.1029/91JC01816.

Minnett, P. J., R. O. Knuteson, F. A. Best, B. J. Osborne, J. A. Hanafin, and O. B. Brown (2001). The Marine-Atmospheric emitted radiance interferometer: A High-Accuracy, seagoing infrared spectroradiometer. *Journal of Atmospheric and Oceanic Technology*, **18** (6), pages 994–1013. doi: 10.1175/1520-0426(2001)018<0994:TMAERI>2.0.CO;2.

Murray, M. J., M. R. Allen, C. J. Merchant, A. R. Harris, and C. J. Donlon (2000). Direct observations of skin-bulk SST variability. *Geophysical Research Letters*, **27** (8), page 1171. doi: 10.1029/1999GL011133.

Nalli, N. R., and L. L. Stowe (2002). Aerosol correction for remotely sensed sea surface temperatures from the national oceanic and atmospheric administration advanced very high resolution radiometer. *Journal of Geophysical Research*, **107** (C10), page 3172. doi: 10.1029/2001JC001162.

Newman, S. M., J. A. Smith, M. D. Glew, S. M. Rogers, and J. P. Taylor (2005). Temperature and salinity dependence of sea surface emissivity in the thermal infrared. *Quarterly Journal of the Royal Meteorological Society*, **131** (610), pages 2539–2557. doi: 10.1256/qj.04.150.

Noyes, E. J., P. J. Minnett, J. J. Remedios, G. K. Corlett, S. A. Good, and D. T. Llewellyn-Jones (2006). The accuracy of the AATSR sea surface temperatures in the caribbean. *Remote Sensing of Environment*, **101** (1), pages 38–51. doi: 10.1016/j.rse.2005.11.011.

O'Carroll, A. G., T. Blackmore, K. Fennig, R. W. Saunders, and S. Millington (2012). Towards a bias correction of the AVHRR pathfinder SST data from 1985 to 1998 using ATSR. *Remote Sensing of Environment*, **116**, pages 118–125. doi: 10.1016/j.rse.2011.05.023.

O'Hara, S., P. Dubock, and J. Abolins (2013). AATSR 12 micron anomaly review board – final report. Technical Report IDEAS-VEG-OQC-REP-1274, Telespazio VEGA, UK.

O'Carroll, A. G., J. R. Eyre, and R. W. Saunders (2008). Three-Way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. *Journal of Atmospheric and Oceanic Technology*, **25** (7), pages 1197–1207. doi: 10.1175/2007JTECHO542.1.

O'Carroll, A. G., J. G. Watts, L. A. Horrocks, R. W. Saunders, and N. A. Rayner (2006). Validation of the AATSR meteo product sea surface temperature. *Journal of Atmospheric and Oceanic Technology*, **23** (5), pages 711–726. doi: 10.1175/JTECH1876.1.

Pinkley, L. W., P. P. Sethna, and D. Williams (1977). Optical constants of water in the infrared: Influence of temperature. *Journal of the Optical Society of America*, **67** (4), pages 494–499. doi: 10.1364/JOSA.67.000494.

Pinkley, L. W., and D. Williams (1976). Optical properties of sea water in the infrared. *Journal of the Optical Society of America*, **66** (6), pages 554–558. doi: 10.1364/JOSA.66.000554.

Rayner, N. A., P. Brohan, D. E. Parker, C. K. Folland, J. J. Kennedy, M. Vanicek, T. J. Ansell, and S. F. B. Tett (2006). Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the Mid-Nineteenth century: The HadSST2 dataset. *Journal of Climate*, **19** (3), pages 446–469. doi: 10.1175/JCLI3637.1.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, **108** (D14). doi: 10.1029/2002JD002670.

Remedios, J. J., R. J. Leigh, A. M. Waterfall, D. P. Moore, H. Sembhi, I. Parkes, J. Greenhough, M. Chipperfield, and D. Hauglustaine (2007). MIPAS reference atmospheres and comparisons to V4.61/V4.62 MIPAS level 2 geophysical data sets. *Atmospheric Chemistry and Physics Discussions*, **7** (4), pages 9973–10017. doi: 10.5194/acpd-7-9973-2007.

Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang (2002). An improved in situ and satellite SST analysis for climate. *Journal of Climate*, **15** (13), pages 1609–1625. doi: 10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.

Roche, A. E., J. B. Kumer, J. L. Mergenthaler, G. A. Ely, W. G. Uplinger, J. F. Potter, T. C. James, and L. W. Sterritt (1993). The cryogenic limb array etalon spectrometer (CLAES) on UARS: experiment description and performance. *Journal of Geophysical Research*, **98** (D6), pages 10763–10775. doi: 10.1029/93JD00800.

Rothman, L., D. Jacquemart, A. Barbe, D. C. Benner, M. Birk, L. Brown, M. Carleer, J. Chackerian, K. Chance, L. Coudert, V. Dana, V. Devi, J. Flaud, R. Gamache, A. Goldman, J. Hartmann, K. Jucks, A. Maki, J. Mandin, S. Massie, J. Orphal, A. Perrin, C. Rinsland, M. Smith, J. Tennyson, R. Tolchenov, R. Toth, J. V. Auwera, P. Varanasi, and G. Wagner (2005). The HITRAN 2004 molecular spectroscopic database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **96** (2), pages 139–204. doi: 10.1016/j.jqsrt.2004.10.008.

Rothman, L. S., A. Barbe, D. C. Benner, L. R. Brown, C. Camy-Peyret, M. R. Carleer, K. Chance, C. Clerbaux, V. Dana, V. M. Devi, A. Fayt, J. M. Flaud, R. R. Gamache, A. Goldman, D. Jacquemart, K. W. Jucks, W. J. Lafferty, J. Y. Mandin, S. T. Massie, V. Nemtchinov, D. A. Newnham, A. Perrin, C. P. Rinsland, J. Schroeder, K. M. Smith, M. A. H. Smith, K. Tang, R. A. Toth, J. V. Auwera, P. Varanasi, and K. Yoshino (2003). The HITRAN molecular spectroscopic database: edition of 2000 including updates through 2001. *Journal of Quantitative Spectroscopy and Radiative Transfer*, **82** (1-4), pages 5–44. doi: 10.1016/S0022-4073(03)00146-8.

Salisbury, J. W., D. M. D'Aria, and F. F. S. Jr. (1993). Thermal infrared remote sensing of crude oil slicks. *Remote Sensing of Environment*, **45** (2), pages 225–231. doi: 10.1016/0034-4257(93)90044-X.

Santee, M. L., A. Lambert, W. G. Read, N. J. Livesey, R. E. Cofield, D. T. Cuddy, W. H. Daffer, B. J. Drouin, L. Froidevaux, R. A. Fuller, R. F. Jarnot, B. W. Knosp, G. L. Manney, V. S. Perun, W. V. Snyder, P. C. Stek, R. P. Thurstans, P. A. Wagner, J. W. Waters, G. Muscari, R. L. de Zafra, J. E. Dibb, D. W. Fahey, P. J. Popp, T. P. Marcy, K. W. Jucks, G. C. Toon, R. A. Stachnik, P. F. Bernath, C. D. Boone, K. A. Walker, J. Urban,

and D. Murtagh (2007). Validation of the aura microwave limb sounder HNO_3 measurements. *Journal of Geophysical Research*, **112** (D24). doi: 10.1029/2007JD008721.

Saunders, P. M. (1967). The temperature at the Ocean-Air interface. *Journal of the Atmospheric Sciences*, **24** (3), pages 269–273. doi: 10.1175/1520-0469(1967)024<0269:TTATOA>2.0.CO;2.

Saunders, R., and P. Brunel (2005). RTTOV 8.7 users guide. Technical Report NWPSAF-MO-UD-008, Met Office, UK.

Schiermeier, Q. (2009). Last weather ship faces closure. *Nature*, **459** (7248), pages 759–759. doi: 10.1038/459759a.

Shenoi, S. C. (1999). On the suitability of global algorithms for the retrieval of SST from the north indian ocean using NOAA/AVHRR data. *International Journal of Remote Sensing*, **20** (1), pages 11–29. doi: 10.1080/014311699213578.

Simmons, A., S. Uppala, D. Dee, and S. Kobayashi (2007). ERA-Interim: new ECMWF reanalysis products from 1989 onwards. ECMWF Newsletter 110, ECMWF, Reading. URL `http://www.ecmwf.int/publications/newsletters/pdf/110_rev.pdf`.

Smith, D. (1999). AATSR infra-red radiometric calibration report. AATSR Project Document PO-TR-RAL-AT-0024, Rutherford Appleton Laboratory.

Smith, D., C. Mutlow, J. Delderfield, B. Watkins, and G. Mason (2012). ATSR infrared radiometric calibration and in-orbit performance. *Remote Sensing of Environment*, **116**, pages 4–16. doi: 10.1016/j.rse.2011.01.027.

Smith, D., R. Watkins, and D. Corney (1993). Test and calibration of the along track scanning radiometer-2 - final report. ATSR-2 Project Documentation ER-RP-OXF-AT-2001, Oxford University.

Smith, D. L., J. Delderfield, D. Drummond, T. Edwards, C. T. Mutlow, P. D. Read, and G. M. Toplis (2001). Calibration of the AATSR instrument. *Advances in Space Research*, **28** (1), pages 31–39. doi: 10.1016/S0273-1177(01)00273-3.

Smith, T. M., and R. W. Reynolds (2004). Improved extended reconstruction of SST (1854–1997). *Journal of Climate*, **17** (12), pages 2466–2477. doi: 10.1175/1520-0442(2004)017<2466:IEROS>2.0.CO;2.

Stark, J. D., C. J. Donlon, M. J. Martin, and M. E. McCulloch (2007). OSTIA : An operational, high resolution, real time, global sea surface temperature analysis system. In *OCEANS 2007 - Europe*, pages 1–4. IEEE, Aberdeen, Scotland, UK. ISBN 978-1-4244-0635-7. doi: 10.1109/OCEANSE.2007.4302251.

Stricker, N., A. Hahne, D. Smith, J. Delderfield, M. Oliver, and T. Edwards (1995). ATSR-2: the evolution in its design from ERS-1 to ERS-2. *ESA Bulletin*, **83**, pages 32–37.

Tisdale, R. T., D. L. Glandorf, M. A. Tolbert, and O. B. Toon (1998). Infrared optical constants of low-temperature H_2SO_4 solutions representative of stratospheric sulfate aerosols. *Journal of Geophysical Research*, **103** (D19), pages 25353–25370. doi: 10.1029/98JD02457.

Trenberth, K. E., P. D. Jones, P. Ambenje, R. Bojariu, D. Easterling, A. Klein Tank, D. Parker, F. Rahimzadeh, J. Renwick, M. Rusticucci, B. Soden, and P. Zhai (2007). Observations: Surface and atmospheric climate change. In Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, editors, *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 235–336. Cambridge University Press, Cambridge, United Kingdom.

Uppala, S. M., P. W. KAllberg, A. J. Simmons, U. Andrae, V. D. C. Bechtold, M. Fiorino, J. K. Gibson, J. Haseler, A. Hernandez, G. A. Kelly, X. Li, K. Onogi, S. Saarinen, N. Sokka, R. P. Allan, E. Andersson, K. Arpe, M. A. Balmaseda, A. C. M. Beljaars, L. V. D. Berg, J. Bidlot, N. Bormann, S. Caires, F. Chevallier, A. Dethof, M. Dragosavac, M. Fisher, M. Fuentes, S. Hagemann, E. Hólm, B. J. Hoskins, L. Isaksen, P. A. E. M. Janssen, R. Jenne, A. P. Mcnally, J. Mahfouf, J. Morcrette, N. A. Rayner, R. W. Saunders, P. Simon, A. Sterl, K. E. Trenberth, A. Untch, D. Vasiljevic, P. Viterbo, and J. Woollen (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, **131** (612), pages 2961–3012. doi: 10.1256/qj.04.176.

Veal, K. L., G. K. Corlett, D. Ghent, D. T. Llewellyn-Jones, and J. J. Remedios (2013). A time series of mean global skin SST anomaly using data from ATSR-2 and AATSR. *Remote Sensing of Environment*, **135**, pages 64–76. doi: 10.1016/j.rse.2013.03.028.

Walton, C. C. (1985). Satellite measurement of sea surface temperature in the presence of volcanic aerosols. *Journal of Climate and Applied Meteorology*, **24** (6), pages 501–507. doi: 10.1175/1520-0450(1985)024<0501:SMOSST>2.0.CO;2.

Walton, C. C. (1988). Nonlinear multichannel algorithms for estimating sea surface temperature with AVHRR satellite data. *Journal of Applied Meteorology*, **27** (2), pages 115–124. doi: 10.1175/1520-0450(1988)027<0115:NMAFES>2.0.CO;2.

Walton, C. C., W. G. Pichel, J. F. Sapper, and D. A. May (1998). The development and operational application of nonlinear algorithms for the measurement of sea surface temperatures with the NOAA polar-orbiting environmental satellites. *Journal of Geophysical Research*, **103** (C12), pages 27999–28012. doi: 10.1029/98JC02370.

Watts, P. D., M. R. Allen, and T. J. Nightingale (1996). Wind speed effects on sea surface emission and reflection for the along track scanning radiometer. *Journal of Atmospheric and Oceanic Technology*, **13** (1), pages 126–141. doi: 10.1175/1520-0426(1996)013<0126:WSEOSS>2.0.CO;2.

Wentz, F. J. (1977). A two-scale scattering model with application to the JONSWAP '75 aircraft microwave scatterometer experiment. NASA contractor report CR-2919, National Aeronautics and Space Administration. URL http://openlibrary.org/books/OL15222182M.

Wimmer, W., I. S. Robinson, and C. J. Donlon (2012). Long-term validation of AATSR SST data products using shipborne radiometry in the bay of biscay and english channel. *Remote Sensing of Environment*, **116**, pages 17–31. doi: 10.1016/j.rse.2011.03.022.

Woodruff, S. D., S. J. Worley, S. J. Lubker, Z. Ji, J. Eric Freeman, D. I. Berry, P. Brohan, E. C. Kent, R. W. Reynolds, S. R. Smith, and C. Wilkinson (2011). ICOADS release 2.5: extensions and enhancements to the surface marine meteorological archive. *International Journal of Climatology*, **31** (7), pages 951–967. doi: 10.1002/joc.2103.

Worley, S. J., S. D. Woodruff, R. W. Reynolds, S. J. Lubker, and N. Lott (2005). ICOADS release 2.1 data and products. *International Journal of Climatology*, **25** (7), pages 823–842. doi: 10.1002/joc.1166.

Wu, X., and W. L. Smith (1997). Emissivity of rough sea surface for 8–13 μm: modeling and verification. *Applied Optics*, **36** (12), page 2609. doi: 10.1364/AO.36.002609.

Závody, A. M., C. T. Mutlow, and D. Llewellyn-Jones (1995). A radiative transfer model for sea surface temperature retrieval for the along-track scanning radiometer. *Journal of Geophysical Research*, **100** (C1), pages 937–952. doi: 10.1029/94JC02170.

Závody, A. M., C. T. Mutlow, and D. T. Llewellyn-Jones (2000). Cloud clearing over the ocean in the processing of data from the Along-Track scanning radiometer (ATSR). *Journal of Atmospheric and Oceanic Technology*, **17** (5), pages 595–615. doi: 10.1175/1520-0426(2000)017<0595:CCOTOI>2.0.CO;2.

Zeng, X., and A. Beljaars (2005). A prognostic scheme of sea surface skin temperature for modeling and data assimilation. *Geophysical Research Letters*, **32** (L14605). doi: 10.1029/2005GL023030.